

\*\*\*\*\*  
W E S R E L L  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu May 20 09:26:24 1999; MasPar time 3339.47 Seconds

Tabular output not generated.

Title: >US-09-099-053-1  
Description: (1-1548) from US09099053.seq  
Perfect Score: 1548  
N.A. Sequence: 1 GCTCGCGGCTCCCATGGCC.....CCAAACGCTCTGGGCTCCAGC 1548  
Comp: CGAGCGCCGAGGATACCG.....GGTTGGAGACCGAGGTCG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl58  
1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
7:em\_in 8:em\_on 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
13:em\_pl 14:em\_ro 15:em\_sts 16:em\_v1  
genbank111  
17:gb\_ba1 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
22:gb\_in2 23:gb\_on 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
28:gb\_pl2 29:gb\_pr1 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_v1

Statistics: Mean 11.063; Variance 5.685; scale 1.946

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	827	53.4	2560	32	MUSSRM Mouse mRNA for srm (ty	0.00e+00
2	825	53.3	2393	32	MUSTSPK Mouse mRNA for tyrosin	0.00e+00
3	191	12.3	2507	29	HSBRK H.sapiens brk mRNA for	2.43e-122
4	191	12.3	2507	25	A42329 Sequence 1 from Patent	2.43e-122
5	180	11.6	3701	24	CHKTKL Chicken tyrosine kinas	1.43e-113
6	176	11.4	1759	24	CHKSRC Chicken c-src gene, co	2.17e-110
7	176	11.4	1982	24	GGIRKA G.gallus yrk mRNA.	2.17e-110
8	177	11.4	3775	32	MAVSCR M.auratus H-19 provira	3.49e-111
9	175	11.3	3110	37	ALDRM144 Rous sarcoma virus (re	1.35e-109
10	175	11.3	3123	32	RESVH19 Hamster H-19 proviral	1.35e-109
11	173	11.2	1653	37	RESVSRC Rous sarcoma virus Pra	5.25e-108
12	174	11.2	4413	37	ALSRCAC Rous sarcoma virus def	8.43e-109
13	174	11.2	4503	37	ASVPR225T Avian Sarcoma Virus PR	8.43e-109

14	173	11.2	9312	37	RESV6	Rous sarcoma virus gen	5.25e-108
15	173	11.2	9392	37	AF033808	Rous sarcoma virus, co	5.25e-108
16	173	11.2	9625	37	ALRCG	Rous sarcoma virus (pr	5.25e-108
17	172	11.1	1759	24	GGCSRC	Gallus gallus gene c-s	3.26e-107
18	172	11.1	2526	24	XHCFYN	Xiphophorus c-fyn (xfy	3.26e-107
19	172	11.1	5188	37	ACGCSRC	Avian sarcoma virus pr	3.26e-107
20	171	11.0	2545	37	RSVPSRC	Duck adapted Rous sarc	2.03e-106
21	171	11.0	9312	37	RSVSEQ	Rous Sarcoma Virus (du	2.03e-106
22	169	10.9	3107	37	REASV5	Avian sarcoma virus sr	7.81e-105
23	169	10.9	3256	37	ALRDA2	Rous sarcoma virus (Sc	7.81e-105
24	167	10.8	1581	37	RSVSR	Avian sarcoma virus SR	3.00e-103
25	167	10.8	1801	37	ALRVSR	Rous sarcoma virus (SC	3.00e-103
26	167	10.8	2241	37	ALRPF60A	Rous sarcoma virus pho	3.00e-103
27	165	10.7	1641	37	RSVSRCHM	Rous sarcoma virus src	1.15e-101
28	163	10.5	1638	37	RSVPPG2V8	Rous sarcoma virus pb6	4.39e-100
29	161	10.4	2293	32	S37068	src (tsuPl)-thermolabi	1.67e-98
30	161	10.4	2293	32	MMU16805	Mus musculus Src-relat	1.67e-98
31	161	10.4	2354	29	HUMFGR	Human fgr proto-oncoge	1.67e-98
32	161	10.4	9317	37	ALRSRDCG	Rous sarcoma virus gen	1.67e-98
33	158	10.2	9396	37	AF052428	Rous sarcoma virus str	3.91e-96
34	153	9.9	2025	37	REFESV	Gardner-Rasheed feline	3.41e-92
35	147	9.5	1190	21	SLSRK3	S.lacustris srk3 mRNA	1.77e-87
36	147	9.5	1253	21	SLSRK2	S.lacustris srk2 mRNA	1.77e-87
37	147	9.5	1529	37	FESVTP1	Feline sarcoma virus T	1.77e-87
38	144	9.3	1926	29	HUMHCKB	Human hemopoietic cell	3.98e-85
39	144	9.3	2015	29	HUMHCKA	Human hemopoietic cell	3.98e-85
40	142	9.2	1629	37	ALRGSRCA	Rous sarcoma virus src	1.47e-83
41	142	9.2	1960	32	NMHCK	Mouse hck gene for tyr	1.47e-83
42	142	9.2	2002	32	MUSBMK	Murine macrophage gene	1.47e-83
43	137	8.9	1589	29	HSU07236	Human mutant lymphocyt	1.19e-79
44	136	8.8	2032	30	HSLCKB	Human lck mRNA for mem	7.15e-79
45	134	8.7	1702	21	SLSRK4	S.lacustris srk4 mRNA	2.59e-77

ALIGNMENTS

1 MUSSRM 2560 bp mRNA ROD 10-FEB-1999  
Mouse mRNA for srm (tyrosine kinase), complete cds.  
D49427  
D49427.1 GI:684971  
srm; tyrosine kinase.  
KEYWORDS  
SOURCE Mus musculus (strain:C57BL/6) adult male thymus cDNA to mRNA.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 2560)  
AUTHORS  
Kawachi Y.  
TITLE  
Direct Submission  
JOURNAL

Submitted (21-FEB-1995) to the DDBJ/EMBL/GenBank databases.  
Yasuhiro Kawachi, Institute of Clinical Medicine, University of  
Tsukuba, Department of Dermatology; 1-1-1, Ten-nodai, Tsukuba city,  
Ibaraki 305, Japan (E-mail:nakauchi@rtcs1.riken.go.jp,  
Tel:0298-53-3128, Fax:0298-53-3217)  
2 (bases 1 to 2560)  
AUTHORS  
Kawachi Y., Nakauchi, H. and Otsuka, F.  
TITLE  
Identification of a novel cDNA clone encoding protein tyrosine  
kinase in murine skin

J. Invest. Dermatol. 21, 533-538 (1995)  
FEATURES  
Location/Qualifiers  
1. 2560  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/dev\_stage="adult"  
/sex="male"  
/tissue\_type="thymus"  
447. 1937  
/codon\_start=1  
/product="srm"  
/protein\_id="BAA08406.1"  
/db\_xref="PID:d1009002"

/db\_xref="pid:g684972"  
/db\_xref="GI:684972"

/translation="MEFLRLRLFLFFWDKIWPADSEEDIPRIQHDNDNPVEQA  
AAVECFPPARLRFALDFARCAEELSVSRDLRYALKKEGDIYFAQLSGPPS  
TSLVPTYLAKATPEPSPDOPFSGISRAOQOLLLSPANAPAFILRPSESSIGY  
GLVRAQAKVCHVRICMAPSGSLVQEGOLFPSDLLAYVKNKLLIONLPLOCP  
QIPLVDEWEPSPSEFVLRKLGSGFGEWEGWELGWSIPVAVKVISADMKLADLK  
ENCAKSLRHLRLHRLAHSGLSEPVITVTELMGKNLQVYLGSGSKALSLPHLLG  
AOKVMSXLEERRVHRDLAARNLVGDLTKVADFGGLARLLKDDVYSPSSGSKI  
PVRWTPAEANRYVFSKSDWSFGLLYEVFTYQCPYEGMTNHTLQIISRGYRLP  
RPVCAPEAVYVLMVECKWGSPEERPTFAILLREKLNAINRRLHLGLT"

polyA\_site 569 a 742 c 699 g 550 t  
BASE COUNT  
ORIGIN

Query Match 53.4%; Score 827; DB 32; Length 2560;

Best Local Similarity 81.3%; Pred. No. 0.00e+00;

Matches 1073; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

Db 621 CTCTTCGGCGGCTCTAGGACTCTCACTGCTCGATGTGCAGAGGAATCAGCGTCAGCGGT 680  
QY 211 CTCTTCCTTGGCTCTATGACTTCACGGCGGGTGTGGCGGGAGCTGAGTGTCCGCGC 270  
Db 681 GGGACAGACTCTAGCGCTCAAGGAGGAGGGGACTACATCTTTGCCAAGGCTCTCT 740  
QY 271 GGGACAGGCTCTGTGCTCGAAGAGGGGGGGGCTACATCTTCGCACGAGGCTTCG 330  
Db 741 GGTCCACCCAGCAGCGGACTAGTTCCTGTCACTACCTTGCCAAAGCTACCCGAGCGG 800  
QY 331 GGCAGCCAGCGCGGCTCTGCTCCATACCCACGCTGGCCAGGCTTCCTGAGAGG 390  
Db 801 CCTCAGACCAACCTTGGTACTTCAGTGGGATCAGCAGGGCTCAGGCCAGCAGTTCCT 860  
QY 391 CTCTCAGACCAACCTTGGTACTTCAGTGGGATCAGCAGGGCTCAGGCCAGCAGTTCCT 450  
Db 861 TTGCTCTCTGCAATGACACAGGGGCTTCCTCATCCGGGCCAGGAAAGCAGATCGGG 920  
QY 451 CTCTCCCAACCCAGCAACCAAGGGGCTTCCTCATCCGGGCCAGGAGCAGCGCTCGG 510  
Db 921 GGTATTCTCTATCAGTCAGGGCCAGGCCAAAGTCTCCCACTACCGCATCGCATGCA 980  
QY 511 GGTACTACTGTGCTGAGTCCGGGCCAGGCCAAGGTCTCCCACTACCGGTCTCCATGCA 570  
Db 981 CCCAGTGGGAGCTCTATCTCAGGAGGGCCAACTTCCCGAGCTGGATGCACTGCTG 1040  
QY 571 GCTGATGCGAGCTCTACCTCAGAGAGGAGGCTCTTCCTCGGGCTGGAGAGTCTGCT 630  
Db 1041 GTTACTACAGACCAACTGGAAGCTGATCCAGAACCTCTGTGTCAGCCCTGCAATCCC 1100  
QY 631 ACCTACTACAGGCGCAACTGGAAGCTGATCCAGAACCCCTCTGTGTCAGCCCTGCA 690  
Db 1101 CAGATACCTTGGTTCAGGAGGAGTGGGAAGCAGCAGCTTCAGAAATTTGCTTCGGA 1160  
QY 691 CAGAAGCCCGGAGGAGGAGTGTGGAGGGCCACACTCCGAATTCGCCCTTGGGAGG 750  
Db 1161 AAGTGGGTGAAGTTTCTTCGGGAGGAGTGTGGGAAGGCTGTGGCTGCTATCCCT 1220  
QY 751 AAGTGGGTGAAGCTTCTTGGGAGGAGTGTGGGAAGGCTGTGGCTGCTATCCCT 810  
Db 1221 GTGGCAGTGAAGGTTATCAATCAGCTGACATGAAGTGGCAGAGCCTCACCAGAGAGAT 1280  
QY 811 GTGGCAGTGAAGGTTATCAATCAGCTGACCAACATGAAGTCACTGACCTGCCAAGAGATC 870  
Db 1281 GAGCAGTGAAGGCTTGAAGGCTGATGAGGCTGATCCGGCTGCAAGCTATATGTTCCCTC 1340  
QY 871 CAGACACTGAAGGCTTGGGAGGAGGCTGATCCGGCTGCAAGGCTGCTGCTGGGC 930  
Db 1341 GGTGAACCTGTGATACCTGTTACTGAACTCATGGCAGGCAACTTCAAGTCTACCTG 1400  
QY 931 GGGAGGCTTGTATACATAGTACAGCAACTCATGCGAAGGGAACCTCAGCGCTTCCTG 990  
Db 1401 GCGAGCTCTCAGGGAAGGCGCTCAGGCTGCCCTCATCTACTGGGATTTGCTGCGCAGTA 1460

QY 991 GGCACCCCGAGGGCGGGCCCTGCGTCTGCCGCCACTCTCTGGGCTTTGCTGCCAGGTG 1050  
Db 1461 GGTGAGGCGATGAGTACCTGAGGAGGCGCGCTGTCTCCACCGGACTTGGTTCGCAG 1520  
QY 1051 GGTGAGGCGATGAGTACCTGAGGAGGCGCGCTGTCTCCACCGGACTTGGTTCGCAG 1110  
Db 1521 AAGTGTGCTGGTGGTGTGATGACCTCAGCTCAGAGGTAGTGTATTTGGCTTGGCAGACTG 1580  
QY 1111 AAGTGTGCTGGTGGTGGTGGTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1170  
Db 1581 CTCAGGAGTGTGCTTACTTCCCAAGCAGTGGCTCCCAAGATCCCTGTCAGTGGAGCGCA 1640  
QY 1171 CTCAGGAGCAGATCTACTCCCGAGCAGCAGCTCCAAGATCCCGTCAAGTGGAGCAGG 1230  
Db 1641 CTTGAGGCTGTCTAATTAACGCTGTCTTTTCCCAAAAGTCAGATGTCTGTGCTTTGGCATC 1700  
QY 1231 CTTGAGGCGGCAATATGCTGTCTTCTCCAGAGATCAGAGCTCTGTGCTTGGCTTGG 1290  
Db 1701 CTGCTGTATGAGTCTTCACTTATGGCAGTGTCCCTATGAAGGAATGACCAACCATGAG 1760  
QY 1291 CTGCTGCAGAGGTTTACCTATGCGCAGTGTCCCTATGAAGGATGACCAACCATGAG 1350  
Db 1761 ACGCTACAGCAGATGCTGATACCGGCTGCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTG 1820  
QY 1351 ACGCTGCAGCAGATCATGCGAGGCTACCGGCTGCCCGCGCGGCTGCTGCTGCTGCTGCTG 1410  
Db 1821 GTCTATGCTCATGCTAGAGTGTGGAAGGCGAGCGCTGAGGAGCGCTGCCACCTTTGGC 1880  
QY 1411 GTCTAGTGTCTCATGCTGAGTGTGCTGAGGAGCAGCGCGGAGGAGCGGCTTGGC 1470  
Db 1881 ATACTGAGGAGAGCTGAATGCCATAAACAGACGCTCCATCTGGGCTCAGCTGACC 1939  
QY 1471 ACGCTGCGGAGAGCTGCACGCCATCCACAGATGCCACCCCTGAGTCTCCTCAGCTGACC 1529

## RESULT 2

LOCUS MUSTSPK 2393 bp mRNA ROD 06-FEB-1999  
DEFINITION Mouse mRNA for tyrosine-specific protein kinase, complete cds.  
ACCESSION D26186  
NID 9529072  
VERSION D26186.1 GI:529072  
KEYWORDS tyrosine-specific protein kinase.  
SOURCE Mus musculus adult lung cDNA to mRNA, clone\_lib:5' stretch.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS Kohmura, N.  
TITLE Direct Submission  
JOURNAL Submitted (24-DEC-1993) to the DBJ/EMBL/GenBank databases. Naohiro Kohmura, National Institute for Physiological Sciences, Lab. Neurobiology & Behavioral Genetics; 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444, Japan (Tel:0564-55-7744, Fax:0564-55-7741)

## REFERENCE

AUTHORS Kohmura, N., Yagi, T., Tomooka, Y., Oyanagi, M., Kominami, R., Yateda, N., Chiba, J., Ikawa, Y., and Aizawa, S.  
TITLE A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption  
JOURNAL Mol. Cell. Biol. 14 (10), 6915-6925 (1994)

## JOURNAL

MEDLINE  
COMMENT Submitted (24-DEC-1993) to DBJ by: Naohiro Kohmura

Department of Neurobiology and Behavioral Genetics National Institute for Physiological Sciences  
Myodaiji, Okazaki 444  
Japan

Phone: 0564-55-7741  
Fax: 0564-55-7744.

## FEATURES

source

Location/Qualifiers  
1. .2393  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="5' stretch"

```

/dev_stage="adult"
/tissue_type="lung"
279..1769
/gene="srm"
279..1769
/gene="srm"
/codon_start=1
/product="tyrosine-specific protein kinase"
/protein_id="BAA05331.1"
/db_xref="PID:d1005873"
/db_xref="PID:g529073"
/db_xref="GI:529073"
/translaton="MEFLKRLTFLSFFWDKIWPASESEDIPRIQGHNDNPVEQA
AAVEPCFPAPRALRLEFDYFARCAEELSVSQDRLYALKKEGDYIFAQRLSGPPS
TGLVPTVYLAKEPSPDQWYFSGISRAQAQOLLSPANAPCAFIRPSESSIGGY
SLSVRAQAKCHVRICAPSGSLVQEGOLFPSLDALLAYVKTNNKLIQNPLOPCIP
QILVQDNERPSERFVFGKLGEGFGEVWEGWLWLSIPVAVKVIKSNMKLADLRF
EIALSLRHERLIRHAICSLGEPYIVITELMGKNLQVTLGSSSEKALSLPHLLGF
ACQVAGMSYLEERRVVRDLAARNVLDGDLCKVADFLGRLKDDVSPSSGSKI
PKWTAPEAAAYRVFSQSDWVSGFGLLYEVFTYQCQPYEGMTNHTLQOISRGRPL
RPVACPAEVVVLVMECKWSPERPTFAILLREKLNAINRRLHLGLT"
repeat_region
1934..2008
polya_signal
2369..2374
polya_site
2393
BASE COUNT 525 a 688 c 655 g 525 t
ORIGIN

```

```

Query Match 53.3%; Score 825; DR 32; Length 2393;
Best Local Similarity 81.3%; Pred. No. 0.00e+00;
Matches 1072; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

Db 453 CTCCTCCGCGGCTCTAGACTTCACTGCTCGATGTCAGAGGAAGTCTGAGGTCAGCGGT 512
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 211 CTCCTCTCGGCTCTATGACTTCAGCGCGGTGTGCGGGAGCTGAGTGTCCGCCGC 270
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 513 GGGGACAGACTTCTACGCCCTCAAGAGGAGGGGACTACATCTTTGCCAAGGCTCTCT 572
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 271 GGGGACAGGCTCTGCGCCTCGAAGAGGGGCGGCTACATCTTCGACGAGGCTTCG 330
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 573 GGTCCACGACGACGAGTCTAGTCTCTACCTACCTTTCGCAAGGCTACCCCGAGCGG 632
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 331 GGCCAGCCGACGCGGCTCGTGCCCATCACCCACGTGGCAAGGCTTCTCTGAGACG 390
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 633 CCCTCAGACCAACTTGGTACTTCAGTGGGATCAGCAGGGCTCAGGCCCGCAGCTGCTC 692
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 391 CTCCTCAGACCAACTTGGTACTTCAGCGGGGTAGTCGGACCGCAGCAGCGCTGCTC 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 693 TTGTCTCTGCAATGACACGAGGCGCTCTCTATCCGCGCCAGCAAGAGCATCGG 752
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 451 CTCCTCCCGACAGACAGAGGCGCTCTCTATCCGCGCCAGCAGCAGCGCTCGG 510
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 753 GGCTATTCTCTATCAGTCAGGCGCCAGCCAAAGTCTCCCACTACCGCATCTGCATGCA 812
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 511 GGCTACTCAGTCTCAGTCCGCGCCAGCCAAAGTCTCCCACTACCGGCTCTCCATGCA 570
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 813 CCAGTGGCAGCTCTATCTCAGAGGGCCAACTTCTCCCGAGCTGGATGCTGCTG 872
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 571 GCTGATGGCAGCTCTACTCAGAAAGGAGGCTCTTTCGCGGCTGGAGGAGTGTCT 630
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 873 GCTTACTACAGACCACTGAGTGCATCAACACCTCTGCTCAGCCCTGCATACCC 932
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 631 ACCTACTACAGGCACTGGAAGCTGATCAGAAACCCCTGCTGCAGCCCTGCATGCC 690
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 933 CAGATACCTTGGTTCAGGAGGAGTGGAAAGCACCCTTCAGATTTGTCTTCGGAAGA 992
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 691 CAGAAGCCCCGAGCAGGAGCTGTGGAGCGGCCACACTCCGAATTCGCCCTTGGAGG 750
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 993 AAGCTGGGTGAAGTCTTCTCGGAGAGTGTGGGAAGCGCTGTGCTGGGCTTATCCCT 1052
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 751 AAGCTGGGTGAAGTCTTCTGGGAGAGTGTGGGAAGCGCTGTGCTGGGCTCCCTGCC 810
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1053 GTGCCAGTGAAGTTATCAATCAGTCACATGAGCTGGCAGACCTCACCAGGAGATT 1112
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

QY 811 GTGGGATCAAGGTCAATCAAGTCAGCCAAACATGAAGTCACTGACCTCCCGAAGGAGATC 870
Db 1113 GAGGCACGAAGAGCTTGAGGCATGAGAGGCTGATCCGGCTGCACGCTATATGTTCCCTC 1172
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 871 CAGACATGAAGGCTCGCGCAGCAGCGGCTCATCCGCTGCACGCACTGTGCTCGGC 930
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1173 GGTGAACCTGTGTACATGCTTACTGAACATCATGGGCAAGGCAACTTGAAGTCTACCTG 1232
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 931 GGGGAGCCTGTGTACATAGTACGGAACATCATCGCAAGGGAACCTGCAGGCTTCTCTG 990
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1233 GGCAGCTCTGAGGGAAGGCCCTGAGCCTGCCCCCTACTTACTGGGATTTGCTTCCAGGTA 1292
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 991 GGCACCCCGAGGGCGGGCCCTCGCTGCTGCGGCCACTCTGGGCTTTGCCCTGCCAGGTG 1050
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1293 GCTGAGGCGATGAGTACCTGCTGAGGAGCGCGCTGCTGCTCCACCGGACATGGCTGCCAGG 1352
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1051 GCTGAGGCGATGAGTACCTGCTGAGGAGCAGCGGCTGTGCTCAGCGGACATTTGGCCGCCCG 1110
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1353 AACGTGCTGTGGTGGTATGACCTACCTGCAAGGTAGTGTATTTGGCCTGGCCAGACTG 1412
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1111 AACGTGCTGTGGAGCGGCTGCTGCTGCAAGGTGGCTGACTTCGGCCTGGCCCGGCTG 1170
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1413 CTCAGGATGATGCTACTCCCAAGCAGTGGCTCCAAAGATCCCTGTCAAGTGGAGGCA 1472
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1171 CTCAGGAGCAGATCTACTCCCGAGCAGCAGCTCAAGATCCCGTCAAGTGGACAGCG 1230
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1473 CTTGAGGCTGCTAATACCGTCTCTTTCCCAAGAGTCAGATGCTGGTCTTGGCATC 1532
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1231 CTTGAGGCGGCAATATGCTGTCTTCTCCAGAACTCAGACGTCTGGTCTTTCGGCGTC 1290
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1533 CTGCTGTATGAGTCTTCACTTATGGCCAGTGTCCCTATGAAGGAATGACCAACCATGAG 1592
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1291 CTGCTGACGAGTCTTCACTATGGCAGTGTCCCTATGAAGGAATGACCAACCATGAG 1350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1593 ACGCTACAGCAGATGCTGCTGATACCGGCTGCCACGCCAGCTGTCTGCCAGCAGAG 1652
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1351 ACGCTGACAGCAGATCATCGAGGCTACCGGCTGCCGCGCCGCTGCTGCCCGGAG 1410
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1653 GTCTATGCTCATGCTGATGCTGGAAGGCGAGCCCTGAGGAGGCTGCCACCTTTGCC 1712
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1411 GTCTAGTGTCTATGCTGAGTGTGAGGAGCAGCGCCGAGGAGCGGCTCTCTTGGC 1470
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1713 ATACTGAGGGAAGCTGAATGCCATACACAGCGCTCCATCTGGGCTCAGCTGACC 1771
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1471 ACGTGGGGAAGCTGCACGCCATCCACAGATGCCACCCCTGAGTCTCAGTGAACC 1529
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 3 HSBK 2507 bp mRNA PRI 19-JUL-1994
LOCUS H.sapiens brk mRNA for tyrosine kinase.
DEFINITION X78549
ACCESSION 9515025
VERSION X78549.1 GI:515025
KEYWORDS brk gene; Tyrosine kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2507)
AUTHORS Mitchell,P.J., Barker,K.T., Martindale,J.E., Kamalati,T.,
Lowe,P.N., Page,M.J., Gusterson,B.A. and Crompton,M.R.
Cloning and characterisation of cDNA encoding a novel non-receptor
tyrosine kinase, brk, expressed in human breast tumours
Oncogene 9 (8), 2383-2390 (1994)
JOURNAL 9430916
MEDLINE 2 (bases 1 to 2507)
REFERENCE Mitchell,P.J.
AUTHORS Direct Submission
TITLE Submitted (28-MAR-1994) P.J. Mitchell, Institute of Cancer
Research, Haddon Labs, 15 Cotswold Road, Sutton, Surrey, SM2 5NG,
UK
FEATURES Location/Qualifiers
source 1..2507

```





Db 751 GCTGGCGCACAACATCTCGGCTGTATACGGCTGTGTCCGTGGGAGACCCCGTGA 810  
QY 885 CTTGGCGCAGAGCGGCTCATCGGCTGCACGAGTGTCTCGGCGGGAGCGCTGTGA 944  
Db 811 CATCATCAGGAGCTCATGCGCAGGAGCGCTGCTGGAGCTGCTCGCGCACTGTATGA 870  
QY 945 CATAGTCACGGAACCTATGCGCAAGGGAACTCTGCAGGCGCTTCTGGGCAACCCCGAGGG 1004  
Db 871 GAAAGTCTCCCTGCTGCGAGCTGTGGAGCTGCTGACATCGCTGCGAGTGGCTCAGGCGCATGTG 930  
QY 1005 CCGGGCCCTGCTGCTCGCCGCACTCTCGGCTTGTGCTGCCAGTGGCTCAGGCGCATGAG 1064  
Db 931 TTACTGGAGTCGCAAAATATACATCCAGCGGAGCTGCGCGCAGGAACATCTCTGTCGG 990  
QY 1065 CTACCTGGAGGAGCAGCGCTTGTGACCGGAGCTTGGCGCGCCGGAACGCTGCTGCTGGA 1124  
Db 991 GGAACACCCCTCTGCAAGTGGGACTTCGGGTAGCGGCTTATCATCAGGAGGACGT 1050  
QY 1125 CGACGGCCTGGCTGCAAGTGGCTGACTTCGGCTTGGCGCGCTGCTCAAGGACGACAT 1184  
Db 1051 CTACCTCTCC--CATGACCAC-AATATCCCTACAACTGGAGTGGCGCCCTGAAGCGCTCTC 1107  
QY 1185 CTACTCCCGAGGAGCAGCTCCCAAGATCCCGGTCAAGTGACAGCGCTCAGGCGGCCAA 1244  
Db 1108 CCGAGGCCATTACTCCACCAATCCGACGCTGCTGCTTTGGGATCTCTGTCATGATGAT 1167  
QY 1245 TTATCTGTCTTCTCCAGAGTCAGACGCTGCTGCTCTCGGCGTCTCTGTCACGAGGT 1304  
Db 1168 GTTCACGAGGGTCAGGTGCGCTTACCGAGCATGTCTCAACCATGAGGCGCTTCTGAGGGT 1227  
QY 1305 TTTCACTATGGCAGTGTCCCTATGAAGGGATGACCAACGAGAGAGCGCTGACGACAT 1364  
Db 1228 GGACGCGGCTACCGATCGCTCGCTCTGGAGTGGCGCCCGCAGCGTGCACAGCTGAT 1287  
QY 1365 CATGCGAGGGTACCGGTGCGCGCGCGGCTGCTGCGCGCGGAGTCTACGTGCTCAT 1424  
Db 1288 GGTGACATGCTGTGAGGAGACCCCGAGCAGACCGCTGCTTCAAGGCGCTTGGGGAGAG 1347  
QY 1425 GCTGGAGTCTGGAGGAGGAGCGCCCGAGGACGCGCCCTCTTGGCCACGCTGCGGGAGAA 1484  
Db 1348 GCT 1350  
QY 1485 GCT 1487

RESULT 5  
LOCUS 3701 bp mRNA VRT 29-JUL-1992  
DEFINITION Chicken tyrosine kinase protooncogene (c-tkl) mRNA, complete cds.  
ACCESSION J03579  
NID 9212712  
VERSION J03579.1 GI:212712  
KEYWORDS c-tkl gene; lck-related proto-oncogene; protein-tyrosine kinase; proto-oncogene; src-related proto-oncogene.  
SOURCE Chicken (strain SPAFAS) spleen, cDNA to mRNA, clone C1.  
ORGANISM Gallus gallus  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 3701)  
AUTHORS Streibhardt, K., Mullins, J. I., Bruck, C. and Ruebsamen-Waigmann, H.  
TITLE Additional member of the protein-tyrosine kinase family: the src- and lck-related protooncogene c-tkl  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84, 8778-8782 (1987)  
MEDLINE 88097370  
FEATURES  
Location/Qualifiers  
1..3701  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
c1..3701  
/note="c-tkl mRNA"  
22..1392  
/note="c-tkl tyrosine kinase (EC 2.7.1.112)"  
/codon\_start=1

/protein\_id="AAA49081.1"  
/db\_xref="PID:g212713"  
/db\_xref="GI:212713"  
/translation="MSPPCSPLQDKLVVLYDYEPHTHGDGLKQKGLRVLRESGEW  
WRAQSUITGQGLPHNFVAMNSLEPEPFNLSKNAEARLNASNTHGSLIRE  
SEYSGSYSLVDFDNOGNETVYKHYKIRMDNGYIYISPRVTFSSHELVEYSSSS  
DGLCTRLGPKRTQKPKWMDQEWEPRESLKLVEKLGAGQFQVWNGFYNGHTKVA  
IKNLKQSGMSPSAFLAEANLMLQHPRLVRLVAVTKEPIYIITEYMEKSLVDFLK  
TSGIKLSINKLLDMAAQIAEGMAFTEAKNYIHRDLRAANILVSEALCECKTADFGLAR  
LIEDNEYTAREGAKFPKWTAPPAINAGTETIKSDVMSFGILLTEIYTGRIYPGMT  
NPVIONLERGYRMPQPDNCPQELYMOCWKEQPERFTFEYMKSVLEDFFTATEG  
QYQQQP"  
BASE COUNT 855 a 1054 c 1089 g 703 t  
ORIGIN  
Query Match 11.6%; Score 180; DB 24; Length 3701;  
Best Local Similarity 63.1%; Pred. No. 1.43e-113;  
Matches 493; Conservative 0; Mismatches 283; Indels 5; Gaps 5;  
Db 546 GCCCGAGAAGCGGTGGGAGGACGAGTGGAGGTGGCAGGAGTCCCTGAAGCTGCT 605  
QY 587 GCCCGAGAAGCGCGGAGGAGGACGTGGGAGGCGGCACACTCCGAATTCGCCCTTGG 746  
Db 606 GGAGAAGCTGGGAGCGCGGCACTTTGGAGAAGTCTTGAATGGGCTTCTACAACGCCACAC 665  
QY 747 GAGGAAGCTGGTGAAGGCTACTTTGGGAGGTGTGGGAAGGCTGTGGCTGGGCTCCT 806  
Db 666 CAAGGTAGCATCAAGAACCTGAAGCAGGCGAGTATGTCCCCAGCGGCTTCTCTGGCGCA 725  
QY 807 GCCCGCGCATCAAGGTCTCAAGTCAAGCAACATCAAGTCACTGACCTCGCCCAAGA 866  
Db 726 GGCCAACTGATGAAGAACCTGACGAGCCACGCGTGGTGGGCTCTACGCTGTG-G-T- 782  
QY 867 GATCAGACACTGAAGGCGCTTGGCGCAGCGCGCTCATCCGGCTGACCGAGTGTGCTC 926  
Db 783 GACCAAGGAGCCATCTACATCATCAGAGATACATGAGAAGGAGCGACCTTGGTGGACTT 842  
QY 927 GGGCGGAGGCTGTGTACATAGTACGGAATCATGCGCAAGGGAACCTGACGAGCCTT 986  
Db 843 CCTCAAGACCTCAGAGGCGCATCAAGCTCAGCATCAACAAATCTTGGACATGCGCGCACA 902  
QY 987 CTTGGCAGCCCGCGGCGGCGCTGCTGCTGCGCGCACTCTCTGGGCTTGTGGCGCA 1046  
Db 903 GATTGCTGAAGGATGCGCTTCACTCAAGCAAGCAACTACATCCACCTGACCTGCGGCGC 962  
QY 1047 GGTGGCTGAGGCGATGAGCTACCTTGAAGGAGCAGCGCGTGTGTCACCGGAGCTTGGCGCG 1106  
Db 963 TGCCCAACATCTCTGTGCGAGGCGCTGTGCTGCAAAATCGCTGCTGGGCTGGCGCG 1022  
QY 1107 CCGGAAGCTGCTGCTGGAGCAGCGGCTTGGCTGCAAGGTGGCTGACTTCGGGCTGGCGCG 1166  
Db 1023 CCTCATCGAGCAACGAATACACAGCAGCAGAGAGGCGCTAAATTCGCCATCAAGTGAC 1082  
QY 1167 GCTGCTCAAGGAGCAGACATCTACTCCCCGAGCAGCAGCTCCAAGATCCCGGTCAGTGGAC 1226  
Db 1083 AGCACCAGGAGCTATCAATACGCGCAGCTTCCACCATCAAGTCTGACGTCTGGTCTCTTGG 1142  
QY 1227 AGCGCTTGAAGGCGCAATATCTGCTTCTCCCAAGAACTCAGACGCTGCTGGTCTCTGG 1286  
Db 1143 CATCTCTCTACTGAGATTGTACTACGCGCGGATCCCGTATCCAGGATGACCAACCC 1202  
QY 1287 CGTCTCTCTGACAGAGTTTTCACCTATGGCCAGTGTCCCTATGAAGGATGACCAACCA 1346  
Db 1203 CGAGGTGATCCAGAACCTGGAGCGCGCTACCGCATGCGCAGCGCCGCAACTGCGCGCA 1262  
QY 1347 CGAGACCTGACGACAGATCATCGAGGATCCGCTGCGCGCGCGGCTGCTGCGCGCGC 1406  
Db 1263 GGAGCTGTACGAACTGATGATGATGCTGGAGGAGGAGCGCT-GAGGAGGCGCGCACCT 1321  
QY 1407 GGAGGTCTACGTGCTCATGCTGGAGTGTGGA-GGAGCAGCGCGCGGAGGAGCGCGCTCT 1465  
Db 1322 T 1322

Qy 1466 T 1466

RESULT 6

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

CHKSRC 1759 bp DNA VRT 19-JUL-1995  
Chicken c-src gene, complete cds (introns deleted).

J00844

9212700

GI:212700

c-myc proto-oncogene; proto-oncogene; src oncogene.

Gallus gallus

Gallus gallus

Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Archosauria; Aves; Neognathae; Galliformes;

Phasianidae; Phasianinae; Gallus.

1 (bases)

Hanafusa, H.

Unpublished (1983)

2 (bases 1 to 1759; 1 to 1759)

Takeya, T. and Hanafusa, H.

Structure and sequence of the cellular gene homologous to the src

src gene and the mechanism for generating the transforming virus

Cell 32 (3), 881-890 (1983)

83155864

JOURNAL

MEDLINE

COMMENT

[2] reports the sequence of c-src without its introns and includes the following table of the predicted splicing junctions based upon the structural similarity between c-src and v-src.

exon

intron

(1) ggagctgac/gtaagcag (990bp) ctccag/ccacacaca  
 (2) gcactggctg/gtagtggtg (50bp) cgtgttcag/gaaaggtgac  
 (3) tcaacacac/gtagtggtg (204bp) tctcttgac/gaaaggtgac  
 (4) aggcgaaga/gtaattgctt (390bp) gctccacag/gtggacttt  
 (5) acgcacaag/gtagtcctgt (1010bp) ataccctag/gtgcctattg  
 (6) tactactca/gtctgtatgg (350bp) actgtctag/aacatcgga  
 (7) gtctggttg/gtaagacag (85bp) tctctcag/ggagcttct  
 (8) ataggaag/gtaggacag (78bp) tctctcag/ggagcttct  
 (9) ggctgcag/gtgggttctt (61bp) cgcctccag/attgcatcg  
 (10) gacggcag/gtggagcgg (118bp) cccgcccag/gtgcacagt  
 (11) ccataccag/gtagagatta (79bp) cttccacag/gtagtgtaa

EMBL features not translated to GenBank features:

key from to description

MSG 22 &gt;1759 messenger RNA

The revision in [1] was a private communication via C. Van Beveren,

19-MAY-1996.

FEATURES

source

gene

CDS

old\_sequence

BASE COUNT

ORIGIN

Location/Qualifiers

1. 1759

/organism="Gallus gallus"

/db\_xref="taxon:9031"

112..1713

/gene="src"

112..1713

/gene="src"

/note="putative"

/codon\_start=1

/protein\_id="AA070194.1"

/db\_xref="pid:g901820"

/db\_xref="gi:901820"

/translation="MGSKSKPKDPSQRRSLRPPDSTHHGFPASQTPNKTAAPDTH

RTPSRGTVAIEPLRGFNSTVTSPPQAGALAGGVTFVALIDYESTETDLSP

KGRQLVNTNEDGWLHSLTTGOTGIPSNVAPSDSQAEWFGKITRRESR

LLNPNRGTFLVRESSTGAYCLSVDFDNAGLNKVKHKIKLSDSGGFIYSRT

QFSLQVLVAYSKHADGLCHRLNVCPTSKPOTGGLAKDAWEIPRESLRLEVLGQ

CFGEWMGTWNGTRVAITLKPMTSPAELEQAOVMKLRHEKLVQLYAVSEPI

YIVTEYKSKLLDFLKGEMKYLRLPQVMAAGASGMAYVEMNTVHRLDRAFI

LVGNLVCKVADFGLIEDNETYARQAKFIKWTAPAAALYKRFITKSDVWSFI

LLTELTGKRPVPGVMNREVLQDQVGYRMPCEPESLHDLMLCQWRDPERPT

FEYLQAFLEDYFTSTPOYPQGENL"

1013

/gene="src"

/citation-[2]

370 a 554 c 533 g 302 t

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

1;

11.4%; Score 176; DB 24; Length 1759;

63.3%; Pred. No. 2,17e-110;

455; Conservative 0; Mismatches 261; Indels 3; Gaps 1;

Db 924 GAAGCTGGGGCAGGCTGCTTTGGAGAGGTGGATGGGACCTGGAACGGCACCACAG 983

Qy 750 GAAGCTGGGTGAAGCTACTTTGGGAGGTGGGAAGGCTGTGGTGGGCTCCCTGCC 809

Db 984 AGTGGCCATAAAGACTTCTGAAGCCCGCCACCATGTCCCCGAGGCGCTCTCTGCAGGAAGC 1043

Qy 810 CTGGCGATCAAGGTCAATCAAGTCACCAACATGAAGCTCACTGACCTGCCAAGGAGAT 869

Db 1044 CCAAGTGATGAAGAAGCTCCGGCATGAGAAGTGGTTCAGCTGTACAGTGTGTGCGGA 1103

Qy 870 CCAGACACTGAAGGCGCTGCGGCACGAGCGGCTCATCGGCTGCACGACGTGTGCTCGG 929

Db 1104 AG--AGCCCATCTACATGCTCACTGAGTACATGACACAGGAGGAGGCTCTCTGATTTCT 1160

Qy 930 CGGGGAGCCTGTGTACATAGTACAGGAACCTCATGCCAAGGGGAACCTGCAGCCTTCT 989

Db 1161 GAAGGAGAGATGGGCAAGTACTCTGGGCTGCCACAGCTCGTCGATATGGCTCTCAGAT 1220

Qy 990 GGGCRCCCGGAGGCGCGGCGCTGCTGCGCCACCTCTCTGGGCTTGGCTTGCCTGCCAGT 1049

Db 1221 TGCATCCGCGATGGCCTATGTGGAGAGATGAACACTAGTGCACCGAGACCTGCGGGCGC 1280

Qy 1050 GGCTGAGGCGATGAGTACCTGGAGGAGCAGCGCTGTGTGCACCGGACTTGGCGCGCG 1109

Db 1281 CAACATCTGTGGGGGAGAACCTGTGTGCAAGTGGCTGACTTTGGGCTGCGACGCT 1340

Qy 1110 GAACCTGCTCGTGGAGCGGCTTGGCCTTGAAGTGGTGTGACTTGGCGCTGCGCGCGCT 1169

Db 1341 CATCGAGGACAAGAGTACACAGCAGCGCAAGTGTCCCAAGTCCCATCAAGTGACAGC 1400

Qy 1170 GCTCAGGACAGACATCTCTCCCGAGAGCAGCTCCAAGATCCCGTCAAGTGGACAGC 1229

Db 1401 CCCGAGCAGCGCCTCTATGGCGGTTTCCACATCAAGTGGATGTCTGTCTTCCGCGAT 1460

Qy 1230 GCTGAGGCGGCCAATATATCGTGTCTTCCAGAGAGTCAAGAGTGTCTGTCTTCCGCG 1289

Db 1461 CTGCTGACTGAGTGAACCAAGGCGGGTGCATACCCAGGATGTCAACAGGGA 1520

Qy 1290 CTTGTGACAGAGTTTTCACCTATGGCAGTGTCTTATGAAGGATGACCAACACGA 1349

Db 1521 GGTGTGGACAGGTGGAGAGGCTTACCGCATGCCCTGCCCGCGAGTCCCCCGAGTC 1580

Qy 1350 GACGCTGACAGCATCATCGAGGTACCGGCTGCCCGCGCGCTGCTTGGCGCGCGGA 1409

Db 1581 GCTGATGACCTCATGTGCGAGTGTGGCGGAGGAGCCTGAGGAGCGGCGCGCTTTG 1639

Qy 1410 GGTCTACGTGCTCATGCTGAGTGTGGAGAGCAGCCCCGAGGACGCGCTTCTTTG 1468

RESULT 7

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

GGYRKA 1982 bp mRNA VRT 22-APR-1993

G.gallus yrk mRNA.

X67786 X68973

963895

GI:63895

X67786.1

yrk gene.

chicken.

Gallus gallus

Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;

Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 1982)

Sudol, M., Greulich, H., Newman, L., Sarkar, A., Sukegawa, J. and

Yamamoto, T.

A novel Yes-related kinase, Yrk, is expressed at elevated levels in

neural and hematopoietic tissues

Oncogene 8 (4), 823-831 (1993)

93205395

2 (bases 1 to 1982)

Sudol, M.







```
promoter      2964..2970
              /note="put. TATA-box" ...
misc_feature  2986..2991
              /note="pot. polyA signal"
repeat_region 2993..3013
              /note="direct repeat 2 (R)"
repeat_region 3092..3097
              /note="direct repeat 1"
source        3098..3123
              /note="hamster DNA"
              /organism="Mesocricetus auratus"
              /db_xref="taxon:10036"
              /haplotype="21-22 hypodiploid"
              /tissue_type="mesenchymal"
              /cell_type="sarcoma cells induced by XC rescued PR-RSV-C"
              /cell_line="H19"
              /clone_lib="lambda NM1149"
              /clone="H19c12.1"
              /clone="H19c12.1"

BASE COUNT   716 a   843 c   936 g   628 t
ORIGIN

Query Match      11.3%; Score 175; DB 32; Length 3123;
Best Local Similarity 62.8%; Pred. No. 1.35e-109;
Matches 465; Conservative 0; Mismatches 272; Indels 3; Gaps 1;

Db 1643 GAAGCTGGGGCAGGCGCTCTTTGGAGAGGCTGTGGATGGGACCTGGAACGGCACCACCAG 1702
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 750 GAAGCTGGGTGAAGGCTACTTTGGGAGGCTGTGGGAAGGCGCTGGCTGGGCTCCCTGCC 809

Db 1703 AGTGCCATAAGACTCTGAAGCCCGGACCATGTGCCCGAGGCGCTTCCCTACAGAAAGC 1762
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 810 CGTGCGATCAAGGTCATCAAGTCAGCACAATGAAGTCACTGACCTCGCCCAAGGAGAT 869

Db 1763 CCAAGTGTGAAGAACTCCGGCATGAAGGCTGTTCAGCTGTACGAGTGGTGTCTCGGA 1822
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 870 CCAGACATGAAGGCGCTCGGCAGCAGCGCTCATCGGCTGCACGAGTGTGCTCGG 929

Db 1823 AG---AGCCCCATCTACATCGCTATTGATATGATAGCAGAGGGAGCCCTTCCTGGATTTCT 1879
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 930 CGGGAGCGCTGTGTACATAGTACACGAACTCATGCGAAGGGGAACTTCGACGGCTTCT 989

Db 1880 GAAGGAGAGATGGCAAGTACCTCGGCTGCCAGCTGCCAGCTCTCGATATGGTGTCTCAGAT 1939
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 990 GGGACACCCGAGGCGCGGCGCTCGCTGCGCCCACTCTCTGGGCTTGGCTGCCAGGT 1049

Db 1940 TGCATCCGGCATGGCTATGTGGAGAGAATGAATACGTGACCGAGAGCCTCGCGGCGGC 1999
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1050 GGCTGAGGCGATGAGCTACTTGGAGGAGCAGCGCGCTTGTGCACCGGAGCTTGGCGCGCG 1109

Db 2000 CAACATCTGTGGGGAGAACCTGGTGTGAAGTGTGCTGACTTTCGGGCTGGCAGCGCT 2059
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1110 GAAGCTGTCTGGACGACGGCTCGGCTGCGCTGCAAGTGGTGTGACTTCGGCTGGCGCGCT 1169

Db 2060 CATCGAGACACAGAGTACACAGCAGCGCAAGGTGCCAAGTTCCTCCATCAAGTGACACAGC 2119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1170 GCTCAAGGACGACATCTACTCCCGAGCAGCAGCTCCCAAGATCCCGGTCAGGTGGACAGC 1229

Db 2120 CCCGAGCAGCGCTCTATGGCGGTTTACCATCAAGTCGGAATGTCTGTCTCTCGGAT 2179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1230 GCCTGAGCGCGCAATATCTGTCTTCTCCAGAAATCAGACGTCGTGGTCTCTCGGCGT 1289

Db 2180 CTGCTGACTGAGCTGACCAACAGGGCGGGTGCCATACCCAGGAGTGGTCAACAGGGA 2239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1290 CTGCTGACGAGGTTTCAACCTATGGCCAGTGTCTCTTATGAAGGATGACCAACACGA 1349

Db 2240 GGTGCTGACGAGTGGAGAGCGGTACCGCATGCCCTGCCCGCCCGAGTGCCTCCGAGTC 2299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1350 GACGCTGACGAGATCATCTCGAGGGTACCGGCTGCCGCGCGGCTGGCTGCCCGCGGA 1409

Db 2300 GCTGATCACTGTGCCAGTGTCTGGGAAGGACCTCGAGGAGCGGCCACCTTTAA 2359
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1410 GGTCTAGTGTCTATCTGGAGTGTGGAGGAGCAGCCCCGAGGAAGCGCCCTCTTTCG 1469
```

```
Db 2360 GTACCTGCAGGCCAGCTGC 2379
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1470 CAGCTGCGGGAGAGAGCTGC 1489

RESULT 11
LOCUS      RERSVSRC      1653 bp      RNA      VRL      01-JUN-1989
DEFINITION Rous sarcome virus Pra src gene (pp60).
ACCESSION  X14718
NID        961714
VERSION    X14718.1 GI:61714
KEYWORDS   src gene.
SOURCE     Rous sarcoma virus.
ORGANISM   Rous sarcoma virus
REFERENCE  1 (bases 1 to 1653)
           Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
AUTHORS    Hackett,P.B.
TITLE      Direct Submission
JOURNAL    Submitted (14-MAR-1989) Hackett P.B., University of Minnesota,
           Department of Genetics and Cell Biology, 250 Biosciences Centre, St
           Paul MN 55108-1095, U S A
REFERENCE  2 (bases 1 to 1653)
           Liu,Z.J. and Hackett,P.B.
TITLE      Sequence variation of the Rous sarcoma virus Pra src gene
JOURNAL    Nucleic Acids Res. 17 (10), 3986 (1989)
MEDLINE    89282411
COMMENT    *map: 7054-8706.
FEATURES   Location/Qualifiers
            source          1..1653
                        /organism="Rous sarcoma virus"
                        /db_xref="taxon:11886"
BASE COUNT 354 a   516 c   500 g   283 t
ORIGIN

Query Match      11.2%; Score 173; DB 37; Length 1653;
Best Local Similarity 62.7%; Pred. No. 5.25e-108;
Matches 464; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

Db 888 GAAGCTGGGGCAGGCGCTCTTTGGAGAGGCTCTGGATGGGACCTGGAAACACACCACCAG 947
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 750 GAAGCTGGGTGAAGGCTACTTTGGGAGGCTGTGGGAAGGCGCTGTGGCTGGGCTCCCTGCC 809

Db 948 AGTGCCATAAGACTCTGAAGCCCGGACCATGTCCCGGAGGCGCTTCTTCGAGGAAGC 1007
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 810 CGTGCGATCAAGGTCATCAAGTCAGCACAATGAAGTCACTGACCTCGCCAAGGAGAT 869

Db 1008 CCAAGTGTGAAGAGCTCCGCGCATGAGAAGCTGTTCAGCTGTACGAGTGTGTCTCGCA 1067.
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 870 CCAGACACTGAAGGCGCTTCGGCAGCGGCTCATCGGCTGCACGAGTGTGCTCGGG 929

Db 1068 AG---AGCCCCATCTACATCGCTATTGAGTACATGACGAAGGGAGCGCTCTCGATTTCCT 1124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 930 CGGGAGCGCTGTGTACATAGTACCGGAACCTCATGCGAAGGGGAACCTGCAGGCGCTTCT 989

Db 1125 GAAGGAGAGATGGCAAGTACCTCGCGGTGCCACAGCTGTCTCGATATGGCTGTCTAGAT 1184
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 990 GGGCACCCCGAGGCGCGGCGCTCGCTGCTGCCGCCACTCTCTCGGCTTTCCTGCGCAGGT 1049

Db 1185 TGCATCCGGCATGGCTATGTGGAGAGAATGAACCTACGTGACCCGAGACCTCGGGCGCGC 1244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1050 GGCTGAGGCGATGAGTACCTTGGAGGAGCAGCGCGCTTGTGCACCGGAGCTTGGCGCGCG 1109

Db 1245 CAACATCTGTGGGGAGAACCTGGTGTGCAAGTGGTGTGACTTTCGGGCTGGCAGCGCT 1304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1110 GAAGCTGTCTGGAGCAGCGGCTGGCTTCGAGGTGGCTGACTTCGGCTGGCGCGCGCT 1169

Db 1305 CATCGAGACACAGAGTACACAGCAGCGCAAGGTGCCAAGTTCCTCCATCAAGTGACAGC 1364
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1170 GCTCAAGGACGACATCTACTCCCGCAGCAGCAGCTCCCAAGATCCCGGTCAAGTGACAGC 1229

Db 1365 CCCGAGCAGCGCGCTCTATGGCGGCTTACCATCAAGTCGAGTGTCTGTGCTCTCGGCAT 1424
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1230 GCCTGAGCGCGCAATATCTGTCTTCTCCAGAAATCAGACGTCGTGGTCTCTCGGCGT 1289
```

		Matches	454;	Conservative	0;	Mismatches	262;	Indels	3;	Gaps	1;
Db	1425	CCTGCTGACTGACCTGACCAACAGGCGCGGTGCCATACACAGGATGCTCAACAGGGA	1484								
QY	1290	CCTGCTGACGAGGTTTACCTATGCGCAGTGTCCCTATGAGGATGACCAACACGGA	1349								
Db	1485	GGTGTGGACAGGTGAGAGGGCTACCGCATGCCCTGCCCGCGAGTGGCCCGAGTC	1544								
QY	1350	GACGCTGCACGACATCATGCGAGGGTACCGCTGCCCGCGCGCTGCCCTGCCCGCGGA	1409								
Db	1545	GTTGCTATGACCTATGCTGCGGAGAGGAGCCCTGAGGAAGCGCCACCTTTAA	1604								
QY	1410	GGTCTACGCTCATGCTGAGTGTGGAGAGCAGCCCGAGGAGCGCCCTCTTTC	1469								
Db	1605	GTACCTGCGGCGACGCTGC	1624								
QY	1470	CACGCTGCGGAGACTGC	1489								
RESULT	12	ALRSRCAC	4413 bp ss-RNA	VRL	28-JUL-1992						
LOCUS		Rous sarcoma virus defective mutant-PR2257 src gene encoding									
DEFINITION		p66-src protein, complete cds.									
ACCESSION		M21526									
NID		9210264									
VERSION		M21526.1	GI:210264								
KEYWORDS		long terminal repeat (LTR); p66-src protein; src oncogene.									
SOURCE		Rous sarcoma virus (strain PR-C, replication-defective mutant PR2257) RNA, passed in quail embryo fibroblast cell line C7.									
ORGANISM		Rous sarcoma virus									
REFERENCE		1 (bases 1 to 4413)									
AUTHORS		Geryk J., Dezelee P., Barnier J.-V. V., Svoboda J., Nelyba J., Karakoz I., Rynditch A.V., Yatsula B.A. and Calothy G.									
TITLE		Transduction of the cellular src gene and 3' adjacent sequences in avian sarcoma virus PR2257									
JOURNAL		J. Virol. 63, 481-492 (1988)									
MEDLINE		89094972									
COMMENT		Draft entry and computer-readable sequence for [1] submitted by P. Dezelee 30-NOV-1988.									
FEATURES		Location/Qualifiers									
source		1..4413									
		/organism="Rous sarcoma virus"									
		/db_xref="taxon:11886"									
LTR		1..335									
		/note="5' long terminal repeat"									
repeat_region		235..255									
		/note="5' LTR R-region"									
CDS		707..2470									
		/note="p66-src protein"									
		/codon_start=1									
		/protein_id="AA42583.1"									
		/db_xref="PID:G210265"									
		/db_xref="GI:210265"									
		/translation="MGSSKSKPKDPSRRSLPPDTHHGGFPASOTPNKTAAPDTH RPSRSFQVATPEKLFQGFNTSDVTSQKALAGGVTTFVALDYRSRETDLSF KAGERLQVNTGEGWLAHSLTGTGIPSNVAPSDSIQAEWIFKTRRESER LLNLRPNVNTGFLVRESETHGTGAYCLSDFDNAKGLNKHQIKRLDGGYITST CFSQQLVYATSKHAGDGLCHRLNVCPTSKPTQGLAKDAWEIPRESLRLEKLGQG CFGEVVMGWTGTTVAIKTLKPGTMSPEAFLOEAQVKKLRHEKLVLYAVVSEPI YIVTEVMSKSLDFLKPGMGKYLRLPOLVDNAQIASGMAYVERMNVYHRLRAANI LVGNELVKVAFGLARLIEDNEYTAQGAQPKFWTAPALYIGRTTKSDVWSFGI LLELTITKGRVYPGMVNRVLDQVGRYMPCEPESLHDLMOCKWKQPEERT FYFLQAFLEDYFTSEPPVPAWRPIGELLLELAPEASLWGTGAWLRAEGRFGEQPOS RMWHGEVSGAPSLIKTVLGH"									
LTR		4079..4413									
		/note="3' long terminal repeat"									
repeat_region		4313..4433									
		/note="3' LTR R-region"									
BASE COUNT		1001 a 1226 c 1269 g 917 t									
ORIGIN		181 bp upstream of EcoRI site.									

RESULT	13	ASVPR225T	4503 bp	RNA	VRL	19-JUL-1993
LOCUS		Avian Sarcoma Virus PR2257T.				
DEFINITION		X51863				
ACCESSION		9394712				
NID		X51863.1	GI:394712			
VERSION		c-src gene; env gene; src protein.				
KEYWORDS		Avian sarcoma virus.				
SOURCE		Avian sarcoma virus				
ORGANISM		Viruses; Retrovirus; Retroviridae; Avian type C retroviruses.				
REFERENCE		1 (bases 1 to 4503)				
AUTHORS		Yatsula B.A., Geryk J., Svoboda J., Rynditch A.V., Calothy G. and Dezelee P.				
TITLE		Evolution and origin of PR2257, a recent c-src transducing virus				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 4503)				
AUTHORS		Zubak S.V.				
TITLE		Direct Submission				
JOURNAL		Submitted (08-MAR-1990) Zubak S., Institute Mol. Biol. & Genet., Academy of Science Ukr SSR, 252627 Kiev, Str Zabolotnogo 130, U S R				
REMARK		revised by [4]				

```
REFERENCE 3 (bases 1 to 4503)
AUTHORS Dezelee P.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1993) Dezelee P., URA 1443 CNRS, Institut
Curie-Biologie Bat. 110, Centre Universitaire Orsay 91405, France
COMMENT On Sep 3, 1993 this sequence version replaced gi:61895.
FEATURES
     Location/Qualifiers
         source
             1..4503
                 /organism="Avian sarcoma virus"
                 /note="Quail sarcoma DNA"
                 /db_xref="taxon:11876"
                 <1..>4503
                 <1..>4503
                 /note="Avian Leukemia Virus leader mRNA"
                 /function="gp37 env gene (partial)"
                 1..335
                 /note="5' LTR"
                 707..2470
                 /standard_name="p66src"
                 /function="protein tyrosine kinase"
                 /codon_start=1
                 /evidence=experimental
                 /product="src protein"
                 /protein_id="CAA36156.1"
                 /db_xref="PID:g394713"
                 /db_xref="GI:394713"
                 /db_xref="SWISS-PROT:P15054"
                 /translation="MGSSKPKDPSSRRSLPPDTHHGGFPASOTPNKTAAPDTH
                 KPSRSFTVATTEKLFEGFNTSDTVPQAGALAGVTFVALYDYESRTDLSF
                 KGRQLQTVNTEGDWLAHSLTGTGTGYPISNVAPSDSIOAEWYFGKTRRESR
                 LLNPNRGTFELVRESETKGAYCLSDFDNAKLVKHVKIKLSDSGFYLTST
                 QFSSLOOLVAYSHAGDLCHRLINVCPTSPQOGLAKDAWEIPRSLRLEVLGQG
                 CFGEVWMTGNTTRFAIKTLKPTGMSPEAFLOEAQVKKLRHEKLVOLYAVVSEEP
                 YIVTEYSKSLDFLALGEMGKYLRLPLQVDMAAIAAGMAYVERMYVHRDLRAANI
                 LVGSLNCKVAGDLALIEDNEYTARQGAKEPIKWTAPAEALYCRFTIKSDYNSFGI
                 LLETLTQGRVYPYGMVNVREVLDOVERGYRMPCEPESLHDLMLCOCWRKDPREPT
                 FEYLOALEDYFTETPEVPAPWRPIGLELLLAPEASLWGTGMLRAEGPFGEPQPS
                 RMWHEVSGAPSLKVLGHP"
         misc_RNA
             2471..3260
                 /function="c-src 3' non coding region"
         LTR
             4168..4503
                 /note="3' LTR"
     BASE COUNT 1015 a 1249 c 1295 g 944 t
     ORIGIN
         11.2%; Score 174; DB 37; Length 4503;
         Best Local Similarity 63.1%; Pred. No. 8.43e-109;
         Matches 454; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

Db 1519 GAAGCTGGGGAGGCGCTGCTTTGGAGAGGTCTGGATGGGGACCTGGACGGCACCACAG 1578
Qy 750 GAAGCTGGGTGAAGCTACTTTGGGAGGTGTGGGAAGCGCTGTGGTGGGCTCCCTGCC 809
Db 1579 AGTGCCATAAGACTCTGAAGCCCGGACCATGTCCCGGAGGCGCTTCCTGCAGGAAGC 1638
Qy 810 CGTGCGCATCAAGTCTCAAGTCAGCAACATGAAGTCACTGACCTCGCCCAAGGAGAT 869
Db 1639 CCAAGTGTATGAAGAAGCTCCGGCATGAGAAGCTGGTTCAGCTGTACGCGAGTGGTTCGA 1698
Qy 870 CCAGACACTGAAGGCGCTTGGCGCAGAGCGGTCTATCGGCTGCACGAGTGTGCTCGGG 929
Db 1699 AG---AGCCCCATCTACATCGTCACTAGTACATGACGAGGGGAGCCCTCGTGGATTCTCT 1755
Qy 930 CGGGAGCGCTGTGTACATGATACACGGAATCATGCGCAAGGGGAACTTCGACGGCCTTCCT 989
Db 1756 GAAGGAGAGATGGCAGTAGTACCTCGGCTGCCACAGCTCTGATATGCTGCTGCATCAT 1815
Qy 990 GGGACACCCCGAGGCGCGGCGCTCGCTCTGCGCCACTCTCTGGCCTTGGCTGCCAGGT 1049
Db 1816 TGCATCGCGCATGGCTTATGTGGAGGAGTGAACCTACGTGACCCGAGACCTCGGCGCGCG 1875
Qy 1050 GCGTAGGAGCATGAGTACCTACCTGGAGGAGCAGCGCGCTTGTGCACCGGACTTGGCGCGCG 1109
```

```
Db 1876 CAACATCCTGGTGGGGAGAACCTGTGTGCAAGGTGGCTGACTTTGGCTGCGCAGCGCT 1935
Qy 1110 GAACGTGCTCTGTGGAGACGGCGCTGCGCTGCAAGGTGGCTGACTTCGGCTGCGCCGCT 1169
Db 1936 CATCGAGGACAAACAGTAGTACACAGCAGCGGCAAGTGCACAAAGTTCCCATCAAGTGCACAGC 1995
Qy 1170 GCTCAAGGACGACATCTACTCCCGGAGCAGCAGTCCCAAGATCCCGGTCAAGTGGACAGC 1229
Db 1996 CCCCAGGAGCGCCCTCTATGCGCGGTTCACCATCAAGTCGGATGCTGTCTCTCGGCAT 2055
Qy 1230 GCGTGAAGCGGCAATATCTGCTCTTCTCCAGAGTGCAGACGTGTGCTCTCTCGCGCT 1289
Db 2056 CCGTGTGACTGAGTGACCAACCAAGCGCGGTGCATATACCCAGGAGTGTCAACAGGGA 2115
Qy 1290 CCGTGTGACGAGGTTTCCCTATGCGCAGTGTCTCTATGAAGGATGACCAACACCA 1349
Db 2116 GGTGCTGACCACTGGAGAGGGGTACCGCATGCGCTCCCGCGCGAGTGGCCCCGAGTC 2175
Qy 1350 GACGCTGCAGCAGATCATCGAGGAGTACCGGCTGCCGCGCGGCTGCTGCCCGCGGA 1409
Db 2176 GCTGATGACCTCATGTGCGCAGTGTGCGGGAAGACCTGAGGAGCGGCCACTTTTG 2234
Qy 1410 GGTCTACGTGCTCATGCTGAGTGTGAGGAGCAGCCCCGAGAACGCGCTCTCTTG 1468

RESULT 14
LOCUS RERSV6 9312 bp DNA 09-FEB-1999
DEFINITION Rous sarcoma virus genome, strain Prague C (Pr-C).
ACCESSION V01197
NID 661695
VERSION V01197.1 GI:61695
KEYWORDS coat protein; endonuclease; oncogene; reverse transcriptase.
SOURCE Rous sarcoma virus.
ORGANISM Rous sarcoma virus.
REFERENCE 1 (bases 1 to 9312)
AUTHORS Schwartz,D.E., Tizard,R. and Gilbert,W.
TITLE Nucleotide sequence of Rous sarcoma virus
JOURNAL Cell 32 (3), 853-869 (1983)
MEDLINE 83155662
FEATURES
     Location/Qualifiers
         1..9312
             /organism="Rous sarcoma virus"
             /db_xref="taxon:11886"
         22..101
             /note="U5 RNA"
         380..904
             /product="p19"
         380..2485
             /codon_start=1
             /product="polyprotein gag"
             /protein_id="CAA24512.1"
             /db_xref="PID:g61696"
             /db_xref="GI:61696"
             /db_xref="SWISS-PROT:P03322"
         /translation="MEAVIKVISSACKTYCGKTSRKKEIGAMLSLLOKEGLMLSPSD
         LYPGSDPITAAALSQAMILGKSLTGLVGLKAAAREEQVTSEQAKFWLIGG
         GRVSPPECIEKATERRIDKGEVGETTVORDAKMAPEETAPKTGTCYHCGTA
         IGCNATASAPPYPYVSGLYPSLAGYGEQGGDDTPPGAQSRPAGHAGQAPGA
         LTNARVRESIASTGPPVAMPVVIKTEGPATPLEKLTITLADITRTKGRSPITM
         AEVALMSSPLLDHVNLMRVILGPAIYALMDAWGVLQTVIAAATRDPRHPANG
         GRGTNLRNLGLADGMVGNPQOALLRPELVITASALQAFREVARLAEPAGPW
         ADIMQGPSEFVDNRLIKAVEGSDLPSPASAPIIDCFQKSPDQQLRTAPST
         LTPGEIHKVLDKQKAPLTDQIAAAMSSAIQPLINAVVNRERDGTGSGGRAGL
         CYTCGSPGHQAQCPKKRSNSRERCOLCNGMHNKQCRKRDGNQGRPGKGLSSG
         PWPGEPPAVSLAMTEHKRPLRVILNTGSHPVKQRSVITALLDSDGADITISE
         EDWPTDWPMEANPQIHGGIGGIPMRKSRDMLIELGVINRDSLERPLLLPFAVAMVR
         GSILGRCLQGLRLNLN"
         911..1096
             /product="reading frame p10"
         1097..1815
             /product="reading frame p27"
         1844..2110
             /product="reading frame p27"
```



/product="reading frame p12"  
 2111..2492  
 /product="reading frame p15"  
 2503..5187  
 /note="pol (reverse transcriptase)"  
 5054..6865  
 /note="reading frame env (envelope protein)"  
 7129..8709  
 /note="transforming gene src"  
 9058..9201  
 /note="U3 RNA"  
 BASE COUNT 2212 a 2359 c 2681 g 2060 t  
 ORIGIN

Query Match 11.2%; Score 173; DB 37; Length 9312;  
 Best Local Similarity 62.7%; Pred. No. 5.25e-108;  
 Matches 464; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

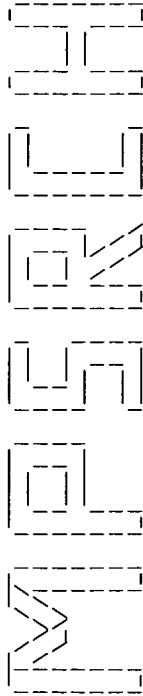
Db	7941	GAAGCTGGGAGGGCTGCTTTGGAGAGGTCTGGATGGGGACCTGGAGACACACACAG	8000
QY	750	GAAGCTGGGTGAAGCTACTTTGGGAGGTGTGGAGAGCGCTGTGGCTCCCTGCC	809
Db	8001	AGTGGCCATAAAGACTCTGAAGCCCGCACCATCTCCCGGAGGCGCTTCTGCAGGAGC	8060
QY	810	CGTGGCCATCAAGTCAATCAAGTCAGGCACATCAAGCTCACTACCTCGCCAGAGAT	869
Db	8061	CCAAAGTATGAAGAAGCTCCGGCATGAGAGCTGGTTCAGCTGTACCGAGTGTGTCGGA	8120
QY	870	CCAGACACTGAAGGGCTCGGCACGAGCGCTCATCGGCTGCACGAGTGTCTCGGG	929
Db	8121	AG---AGCCCATCATCATGCTATTAGTACATAGACAAGGGAGCGCTCTGGAATTCCT	8177
QY	930	CGGGAGCGCTGTACATAGTACGGAAGTACATCGCAAGGGAACCTGCAGGCGCTTCT	989
Db	8178	GAAGGGAGAGATGGCAAGTACCTGGCGCTGCCACACTGCTGCATATGCTGCTCAGAT	8237
QY	990	GGGACCCCGGAGGCGCGGCGCTGCTGCTGCGCCCACTCTGCGGCTTTGCCGCCAGGT	1049
Db	8238	TGCATCCGGCATGCGCTTATGTGGAGAGAATGAACCTAGCTGCACCGAGACCTGCGGGCGGC	8297
QY	1050	GGCTGAGGGCATGAGCTACCTGGAGGAGCAGCGGTTGTGCACCGGAGACTTGGCGCGCG	1109
Db	8298	CAACATCCTGGTGGGGAGAACCTGGTGTGCAAGGTGGCTGACTTCGGGTGGCACCGCT	8357
QY	1110	GAAGCTGCTGTGACGACGCGCTGGCTGCAAGGTGGCTGACTTCGGCTGGCGCGCT	1169
Db	8358	CATCGAGGACAAGAGTACACAGACGCGCAAGTGCACCAAGTTCCTCATCAAGTGGACAGC	8417
QY	1170	GCTCAAGGAGACATCTACTCCCGGACGACAGCTCCCAAGATCCCGTCAAGTGGACAGC	1229
Db	8418	CCCCGAGGACGCCCTCTATGGCGGTTTCAACATCAAGCTCGGATGCTGTGCTTCGGCAT	8477
QY	1230	GCCTGAGCGGCCAATATTCTGCTCTTCTCCAGAACTCAGACGCTGTGCTTCGGCT	1289
Db	8478	CTGCTGACTGAGCTACCAACACCAAGGCGCGGTGCCATACCCAGGAGTGTCTCAACAGGA	8537
QY	1290	CCTGCTGCAGGAGTTTTCACCTATGCGCAGTGTCCCTATGAAGGGATGATCAACACCA	1349
Db	8538	GTGCTGACGAGGTGAGAGGGGCTACCGATGCCCTGCCCGCCAGTGCCTCCCGAGTC	8597
QY	1350	GACGCTGCAGCATATCATGAGAGGGTACCAGGCTGCGCGCGCGGCTCCCTCCCGCGGA	1409
Db	8598	GCTGCATGACTCATGTGCGCAGTGTCTGGCGGAGGAGCCCTGAGGAGGCGGCCACTTTAA	8657
QY	1410	GGTCTAGTGTCTATGCTGAGTGTCTGGAGGAGCAGCCCGGAGGAGGAGGAGGAGGAGG	1469
Db	8658	GTACCTGCAGGCCAGCTGC 8677	
QY	1470	CACGCTCGGGGAGGAGCTGC 1489	

RESULT 15 AF033808 9392 bp RNA VRL 03-DEC-1998  
 LOCUS



Search completed: Thu May 20 13:26:50 1999  
Job time : 14426 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu May 20 14:58:36 1999; MasPar time 400.17 Seconds  
Tabular output not generated. ... 8928:892-Million-cell-updates/sec

Title: >US-09-099-053-1  
Description: (1-1548) from US0909053.seq  
Perfect Score: 1548  
N.A. Sequence: 1 GCTCGGGGCTCCCATGGCC.....CCAACGCTCTGGCTCCAGC 1548  
Comp: CGAGGCGCCGAGGTACCG.....GGTGGAGACCGAGGTGC

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dhase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 9.225; Variance 5.916; scale 1.559

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1548	100.0	1548	59	Human SAD encoding CD	0.00e+00
2	191	12.3	2507	13	Breast tumour kinase,	1.95e-101
3	178	11.5	1602	7	Chicken pp60 c-src ge	4.84e-93
4	153	9.9	1611	7	Human pp60 c-src gene	5.48e-77
5	130	8.4	1254	2	Lck gene fused with p	2.36e-62
6	90	5.8	2820	28	gD, trkA fusion useful	2.17e-37
7	88	5.7	2301	40	Human c-trk oncogene.	3.65e-36
8	81	5.2	5520	29	c-abl gene.	6.88e-32
9	79	5.1	3623	40	Human c-abl oncogene.	1.08e-30

10	79	5.1	3780	29	T61864	c-abl gene.	1.08e-30
11	76	4.9	755	5	Q30725	Potential tyrosine ki	6.98e-29
12	76	4.9	755	54	V70224	Rat orphan tyrosine k	6.98e-29
13	76	4.9	4092	54	V70230	Human receptor tyrosi	6.98e-29
14	68	4.4	1942	14	Q84888	DNA encoding cytoplsm	4.18e-24
15	68	4.4	1987	46	V44497	Human matk cDNA.	4.18e-24
16	68	4.4	2000	17	T00616	Megakaryocyte kinase	4.18e-24
17	63	4.1	1739	4	Q27539	Tyrosine Kinase recep	3.70e-21
18	58	3.7	4049	15	Q90660	Eph-related PTK Cdk5.	3.02e-18
19	58	3.7	4097	15	Q90657	Eph-related PTK Cdk5+	3.02e-18
20	52	3.4	1509	9	Q53470	PTK gene Bl.	8.25e-15
21	52	3.4	2962	17	T02946	EPH-like receptor pro	8.25e-15
22	53	3.4	3453	13	Q79068	Mouse flk-2 cDNA.	2.23e-15
23	53	3.4	3453	9	Q53502	Murine flk-2 cDNA.	2.23e-15
24	53	3.4	3453	13	Q81012	Flk2 receptor protein	2.23e-15
25	53	3.4	3453	7	Q40914	Murine flk-2 cDNA.	2.23e-15
26	53	3.4	3453	6	Q35249	Murine flk-2 coding s	2.23e-15
27	53	3.4	3453	30	T72118	Murine flk-2 receptor	2.23e-15
28	53	3.4	3453	23	T38733	Human foetal liver k	2.23e-15
29	53	3.4	3521	16	T00801	Flk2/flt3 tyrosine ki	2.23e-15
30	52	3.4	3546	15	Q90656	Eph-related PTK Cdk10	8.25e-15
31	52	3.4	3591	15	Q90658	Eph-related PTK Cdk10	8.25e-15
32	52	3.4	3713	7	Q47065	tie truncated recepto	2.23e-15
33	53	3.4	3768	24	T40739	HER-2/neu oncogene.	2.23e-15
34	53	3.4	3845	7	Q47064	tie receptor kinase c	2.23e-15
35	52	3.4	3955	17	T01590	Rat neu promoter.	8.25e-15
36	52	3.4	4281	9	Q53471	elk cDNA.	8.25e-15
37	53	3.4	4299	8	Q46083	Sequence encoding a c	2.23e-15
38	53	3.4	4530	36	T71253	Human HER2 gene.	2.23e-15
39	53	3.4	4530	17	T01585	Her-2/neu (ERBB2)/c-er	2.23e-15
40	51	3.3	3060	28	T51457	gD, trkB fusion useful	3.04e-14
41	51	3.3	3194	20	T00689	Human trkB receptor D	3.04e-14
42	51	3.3	3453	5	Q29954	Murine flk-2 cDNA seq	3.04e-14
43	51	3.3	3663	32	T72320	Embryonic stem cell k	3.04e-14
44	51	3.3	3707	18	Q99277	Human neurotrophic fa	3.04e-14
45	50	3.2	3751	11	Q62461	Human embryonal Kinas	1.11e-13

ALIGNMENTS

RESULT 1	ID	V81743 standard; CDNA; 1548 BP.
AC	V81743;	AC
DT	10-MAR-1999	(first entry)
DE	Human SAD encoding cDNA.	
KW	PTP04; PTP03; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;	
KW	type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;	
KW	neurodegenerative disease; neuronal survival; Alzheimer's disease;	
KW	Parkinson's disease; Huntington's disease; ss.	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	CDS	49..1515
FT	FT	/*tag= a
FT	PT	WO9849317-A2.
PD	PD	05-NOV-1998.
PF	PF	23-APR-1998; US08439.
PR	PR	23-OCT-1997; US-063595.
PR	PR	28-APR-1997; US-044428.
PR	PR	20-MAY-1997; US-047222.
PR	PR	11-JUN-1997; US-049477.
PR	PR	11-JUN-1997; US-049756.
PR	PR	18-JUN-1997; US-049914.
PA	(SUGEN) SUGEN INC.	
PI	App H; Clary D; Courtneidge SA; Hui TH; Jallal B,	
PI	Markby D; Onrust S; Peles E; Plozman GD;	
DR	WPI; 99-009434/01.	
DR	P-PSDB; W89248.	
PT	New nucleic acid encoding specific protein tyrosine phosphatases -	
PT	useful for identifying specific modulators for treatment and	
PT	prevention of cancer and neurodegenerative disease	
PS	Claim 2; Page 146; 199pp: English.	
CC	The present invention describes isolated, enriched or purified nucleic	
CC	acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The	

present sequence encodes human SAD. The above proteins, other than ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify substances that modulate their activity (i.e. agonists and antagonists, including NBP) in vivo or in vitro. These substances are used to treat or prevent diseases associated with abnormal signal transduction pathways that involve the proteins, particularly cancer (e.g. leukemia and lymphoma), while modulators of ALK-7 (which is a type I receptor serine/threonine kinase) are used to promote neuronal survival, particularly for treating Alzheimer's, Parkinson's or Huntington's diseases. Nucleic acid fragments of the polynucleotides encoding the proteins can be used as probes to identify and clone related sequences; to detect protein-encoded RNA; to generate transgenic animals and in gene therapy (optionally after mutation). Ab are used to determine the proteins.

Sequence 1548 BP; 258 A; 533 C; 494 G; 263 T;

Query Match 100.0%; Score 1548; DB 59; Length 1548;

Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 1548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gctcggggctcccatgcccctcggcccgccagcgtgtgaccccggggagtgagccgttc 60  
 Qy 1 GCTCGGGGCTCCCATGCCCCCTCGGGCCCGAGCGTGTGACCCCGGGGATGAGCGGTTC 60  
 Db 61 ctacaggaggcgtggtct 120  
 Qy 61 CTCAGAGGCGGCTGGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120  
 Db 121 gagccggaccatggcaccccggttccctggaccaccaactgaccagtgccacgcttc 180  
 Qy 121 GAGCCGGACCATGGCACCCCGGGTCCCTGGACCCCAACACTGACCCAGTGGCCAGCTC 180  
 Db 181 cccgcggagccttgagcccttccctcagctctctctctctctctctctctctctctct 240  
 Qy 181 CCGCGCGAGCCTTGACGCGCCCTTCCCTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 Db 241 cgggtgcccgggagctagtgatcgcgcgggggacagcgtctgtccctcaagagggg 300  
 Qy 241 CGGTGCGGGGAGCTGAGTGTCCGCGCGGGGACAGGCTCTGTGCCCTCGAAGAGGGG 300  
 Db 301 ggcggtacatcttcgacgagcttctcggccagccagccgagcgtcgtgcccctc 360  
 Qy 301 GCGGCTACATCTTCGACGAGCTTTTCGGGCCAGCCCGAGCCGCGGCTGTGCCCATC 360  
 Db 361 acccagtgggcaggcttctctctgagcgtctctcagaccacccctgttactttagcgg 420  
 Qy 361 ACCCAGTGGCCAAAGCTTCTCTGAGAGCTCTCAGACCAACCTGTGTACTTTAGCGGG 420  
 Db 421 gtcagtggaccagcagcagcagcgtctctctctctctctctctctctctctctctct 480  
 Qy 421 GTACGTGCGGACCCAGGACAGCAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480  
 Db 481 ctcatccggccagcagcagcagcgtcggggctctactactcagtcggtccgcccagggc 540  
 Qy 481 CTCATCCGGCCAGCAGGAGCAGCCTCGGGGGCTACTACTGTACGTCCGGGCCCCAGGCC 540  
 Db 541 aaggtctgccactaccgggtctccatggcagcgtgatggcagcgtctactctcagaaagg 600  
 Qy 541 AAGGTCTGCCACTACCGGCTCTCCATGGCAGCTGATGGCAGCCTCTACCTGCAGAAGGGA 600  
 Db 601 cggctcttcccgccctggagagcgtctcactactactacaaaggcccaactggagctgac 660  
 Qy 601 CGGCTCTTCCCGGCTGGAGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 660  
 Db 661 cagaacccctctcagccttgcagtcgcccagaggcccccagggcagcagcgtgtggag 720  
 Qy 661 CAGAACCCTCTCTCGACCCCTTCATGTGCCCCCAGAGCCCCCAGGAGGACCTGTGGGAG 720  
 Db 721 cggccacactccgaattcggcccttgggaggaagctgggtgaaggctacttttgggaggtg 780  
 Qy 721 CGGCCACACTCCGAATTCGCCCTTGGGAGGAAGCTGGGTGAAGGTACTTTTGGGAGGNG 780  
 Db 781 tgggaaggcctgtgctgggtccctctgcccgtggcgatcaaggtcatcagtcagccac 840

Qy 781 TGGGAAGGCTGTGGCTGGGCTCCCTGCCGTGGCGATCAAGGTCATCAAGTCAGCCAAC 840  
 Db 841 atgaagctcaactgacctcgcgaaggagatccagacactgaagggcctggggcagcagcgg 900  
 Qy 841 ATGAAGCTCACTGACCTCGCAAGGAGATCCAGACACTGAAGGGCCTGCGGCACGAGCGG 900  
 Db 901 ctcatccggtgcacgcagctgtctcggcggggagccctgtgtacatagtcacaggaaactc 960  
 Qy 901 CTCATCCGGCTGCACGACGTGTCTCGGGCGGGAGCCTGTGTACATAGTFCACGGAACTC 960  
 Db 961 atgcccgaagggaacctgcagccttccctgggcaccccgagggccggccctgcgctctg 1020  
 Qy 961 ATGCCGAAGGGAACTGCAGGCTTCTTGGGCACCCCGAGGCGCGGCCCTTGGCTCTG 1020  
 Db 1021 ccgccaactcctgggtttgctgcgcagggtggctgagggcatgagctacctggggagcag 1080  
 Qy 1021 CCGCCACTCTGGGCTTTCCTGCCAGGTGGCTGAGGCGATGAGTACCTTGGAGGAGCAG 1080  
 Db 1081 cgcgttgtcacgcgggaacttggccgcccgggaacgtcgtcgtggacagcggcctggcctgc 1140  
 Qy 1081 CCGCTGTGCACCGGGACTTGGCCGCCCGGAACGCTGCTGTGGACAGCGGCTTGGCTGC 1140  
 Db 1141 aaggtggtgacttcggtcctggtccggtgctcaaggacagacatctactcccgagcagc 1200  
 Qy 1141 AAGGTGCTGACTTTCGGCTTGGCCGCTGCTCAAGGACGACATCTACTCCCCGAGCAGC 1200  
 Db 1201 agctccaaagtccccggtcaagtggagcagcgcgtgagcgcgccaattatcgtctctctcc 1260  
 Qy 1201 AGCTCCAAAGTCCCGGTCAAGTGGACAGCGCTGAGCGGCCCAATTTATCGTGTCTCTCC 1260  
 Db 1261 cagaagtcagaagcgtctggttctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1320  
 Qy 1261 CAGAAAGTCAGAGCTGTGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 Db 1321 tctccctgaaggatgacaaacacagagcgtcagcagatcatcgagggtaccgg 1380  
 Qy 1321 TGTCCCTATGAAGGGATGACCAACACGAGAGCGCTCAGCAGATCATGCGAGGGTACCGG 1380  
 Db 1381 ctgc 1440  
 Qy 1381 CTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
 Db 1441 agcagcccgaggagcggcctctcttggcagcgtcgggagagcgtcgggagagcgtcagccatccac 1500  
 Qy 1441 AGCAGCCCGAGGAGCAGCGCCCTCTTTGCCACGCTCGGGGAGAGCTGCACGCCATCCAC 1500  
 Db 1501 agatgccacccctgagtcctcagtgacccacgctctgggtccacg 1548  
 Qy 1501 AGATGCCACCCCTGAGTCTCTCAGTGACCCAAAGCTCTGGGCTCCAGC 1548

# RESULT 2

ID Q81189 standard; cDNA; 2507 BP.

AC Q81189;

DT 12-AUG-1995 (first entry)

DE Breast tumor kinase, brk.

KW Breast tumor kinase; brk; protein-tyrosine-kinase; PTK;

KW breast cancer; metastasis; prognosis; diagnosis; T-47D; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds 29..1384

FT /\*tag= a

PN W09502057-A.

PD 19-JAN-1995.

PF 08-JUL-1994; G01479.

PR 09-JUL-1993; GB-014233.

PR 11-MAR-1994; GB-004817.

PA (CANC-) CANCER RES INST.

PA (WELL) WELLCOME FOUND LTD.

PI Barker KT, Crompton MR, Gusterson BA, Kamalati T;

PI Mitchell RJ, Page MJ, Spence P;

DR WPI; 95-066901/09.



```
QY 1290 CCTGCTGACGAGGTTTTCACCTATGCGCAGTGTCCCTATGAAGGATGACCAACACGA 1349
Db 1410 ggtctgagaccagtgagaggggtaccgatgccctgcccgcgcagtgcccagtc 1469
QY 1350 GACGCTGACGAGATCATGCGAGGGTACCGCTGCGCGCCCGCTGCTGCCGGCGGA 1409
Db 1470 gctgcatgacctatgtgcccagtgctggcggaggaccctgaggagcgccacctttg 1528
QY 1410 GGTCTAGTCTCATGCTGGAGTGTCTGGAGGAGCAGCCCCCGAGGAACGGCCCTCCTTTG 1468

RESULT
ID Q46688 standard; cDNA to mRNA; 1611 BP.
AC Q46688;
DE 23-DEC-1993 (first entry)
KW Human pp60 c-src gene.
OS Endothelial; tyrosine kinase protein; pp60 c-src; ss.
PN Homo sapien.
PD WO9314193-A.
PD 22-JUL-1993.
PF US004445.
PR 03-JAN-1993; US-820011.
PR 06-JAN-1992; US-820011.
PA (UYUA ) UNIV YALE.
PI Bell L, Luthringer DJ, Madri JA, Warren SL;
DR P-PSDB; R39705.
PT Genetically engineered endothelial cells - which exhibit enhanced
PT cell migration, urokinase-type plasminogen activator activity,
PT and reduced mononuclear cell adhesion and fibronectin prodn
PS Disclosure: Page 69-72; 91pp; English.
CC The DNA encoding a portion or (more preferably) the entire pp60
CC c-src polypeptide is used to transform endothelial cells.
CC Transformed cells produce increased amounts of pp60 c-src and have
CC improved therapeutic properties. They migrate at faster rates than
CC non-transformed counterparts; have an enhanced ability to inhibit
CC the formation of thrombi and/or dissolve thrombi once they have
CC formed and exhibit reduced mononuclear cell adhesion. They can also
CC be used to improve the success of surgical procedures such as
CC coronary angioplasty, heart bypass surgery, vessel graft and stent
CC implantation.
SQ Sequence 1611 BP; 334 A; 507 C; 504 G; 266 T;

Query Match 9.9%; Score 153; DB 7; Length 1611;
Best Local Similarity 61.7%; Pred. No. 5.48e-77;
Matches 442; Conservative 0; Mismatches 271; Indels 3; Gaps 1;

Db 823 aagctggggcagggtcttggcgaggtgtggatgggacctggaacggtaccaccagg 882
QY 751 AAGCTGGGTGAAGCTACTTTGGGGAGGTGTGGGAAGCCCTGTGGCTGCCCTCCCTGCC 810
Db 883 gtggccatcaaaacctgaagcctggcagcatgtctccagaggccttctgcaggaggcc 942
QY 811 GTGGCGATCAAGGTCATCAAGTCAGCCAAACATGAAGCTCACTGACCTGCCCAAGAGATC 870
Db 943 caggtcatgaagaagctgagggatgagaagctggtgagttgtatgtgtgtttcaga- 1001
QY 871 CAGACATGAAGGGCCCTGCGGCACGAGCGCTCATCGGGCTGCACGAGTGTGCTCGGGC 930
Db 1002 --ggagccattacatcgtcacggagtacatagcaaggaggttgcgtgaattctc 1059
QY 931 GGGAGAGCTGTGTACATAGTACAGGAACCTATGCGCAAGGGGAACCTGCGAGCCCTCCTG 990
Db 1060 aagggggagacaggaagtacgtcggtcgctcagctagctggtggacatggctgcagatc 1119
QY 991 GGCACCCCCGAGGGCCGGCCCTGCTGCTGCCGCACTCTCTGGGCTTGTCTGCGAGGTG 1050
Db 1120 gctcaggcatggcgtagctgagggcgatgaactacatccaccgggaccttcgtgcagcc 1179
QY 1051 GCTGAGGGCATGAGTACCTACCTGGAGGACGAGCGCTTGTGCACCGGACTTGGCCGCCCGG 1110
Db 1180 aacatcctggtggagagaacctggtgtgcaaatggtccagacttgggtcgtcgtc 1239
QY 1111 AACGTGCTGCTGAGCAGCGCCCTGGGCTGCAAGGTGCTGACTTCGGCCCTGGCCCGCTG 1170
```

```
Db 1240 attgaagacaatgagtacacggcggcgaagtgccaaattccccatcaatggcagcgt 1299
QY 1171 CTCAAGGACGACATCTACTCCCGGACGACGACTCCAAGATCCGGTCAAGTGCACAGCG 1230
Db 1300 ccagaagctgccccttatggcgcgttccaccatcaagtcgagagtggtgcttcctcgagtc 1359
QY 1231 CCTGAGCGGGCCAATTATCGTGTCTTCTCCCAAGATCAGACGCTCTGGTCTCTCGCGCTC 1290
Db 1360 ctgctgactgagctcaccacaaaggaggggtgcccctaccctgggagtggtgaaccgcgag 1419
QY 1291 CTGCTGACGAGGTTTTCACCTATGCCAGTGTCTCTATGAAGGATGACCAACACGAG 1350
Db 1420 gtgctggaccaggtggagcgggggtaccggatgcccctgcccggagtgctcccgagtc 1479
QY 1351 ACGCTGCAGCAGATCATGCGAGGTTACCGGCTGCCGCGCCGCTGCTGCGCGGAG 1410
Db 1480 ctgcaagacctcatgtgccagtgctggcgggaaggagcctgagggcgccaccctt 1535
QY 1411 GTCTACGTGCTCATGCTGGAGTGTGGAGGACGACGCCCGGAGAACGGCCCTCTT 1466

RESULT
ID Q13983 standard; DNA; 1254 BP.
AC Q13983;
DE 13-DEC-1991 (first entry)
KW Lck gene fused with part of beta-galactosidase gene.
OS Synthetic.
PE Key Location/Qualifiers
FT misc_rna 1..78
FT /tag= a
FT /note= "beta-galactosidase gene fragment"
FT misc_rna 79..1254
FT /tag= b
FT /note= "lck gene"
FN J03201994-A.
PD 03-SEP-1991.
PF 28-DEC-1989; 338268.
PR 28-DEC-1989; JP-338268.
PA (TOKU ) TOKUYAMA SODA KK.
DR WPI; 91-300980/41.
DR P-PSDB; R14201.
PT Fused polypeptide - has amino acid sequence of beta-galactosidase
PT with a LCK gene conjugated to the N-terminal via DNA having
PT multi-cloning site
PS Disclosure; Fig 4.2; 15pp; Japanese.
CC The sequence consists of the first 78 bp encoding the N-terminal
CC amino acids of the beta-galactosidase gene fused with the lck gene.
CC It is prepd. by a claimed process in which a DNA contg. the lck
CC gene is inserted into an E.coli expression vector. The vector has
CC DNA contg. part or all of the beta-galactosidase gene at the
CC appropriate site of the multi-cloning site. It is useful for
CC producing an antibody specifically immunoreactive with only a lck
CC gene-derived polypeptide in T cells. The antibody may recognise
CC lck gene-derived polypeptides in human cells.
SQ Sequence 1254 BP; 291 A; 361 C; 365 G; 237 T;
```

```
Query Match 8.4%; Score 130; DB 2; Length 1254;
Best Local Similarity 62.1%; Pred. No. 2.36e-62;
Matches 373; Conservative 0; Mismatches 225; Indels 3; Gaps 3;

Db 599 tgaagcagctcaacacacagcggctggttcggtcctacgtctg-g-lcaccacagg-agc 655
QY 878 TGAAGGGCTCGGGACAGCGGCTCATCCGGCTGCACGAGTGTCTCGGGCGGGGAGC 937
Db 656 ccattcatcatcactgaatacataggagaatggagcttagtgatttctcaagacc 715
QY 938 CTGTGTACATAGTCACGAACTATGCGCAAGGGGAACCTGCAGGCCTTCCTGGGCACCC 997
Db 716 ctccaggcatcaagttgaccatcaacaaactcctggacatggcagcccaaatgagaag 775
QY 998 CCGAGGGCGGGCCCTCGCTGCTGCCGCTTCCCTGGGCTTTGCTGCCAGTGGCTGAGG 1057
```







PF 15-SEP-1994; 306691.  
PR 15-SEP-1994; US-306691.  
PA (UJTF-) UNIV JEFFERSON THOMAS.  
PI Calabretta B, Skorski T;  
PI WPI; 98-229882/20.  
PT Anticancer composition comprising two anti-sense oligo:nucleotide(s)  
PT - targeting cytoplasmic and nuclear oncogene(s)  
PS Claim 1; Column 105-110; 92pp; English.  
CC The present invention describes an oncogene from the present invention.  
CC The present invention describes a composition which comprises two  
CC antisense oligonucleotides. The first oligonucleotide is specific for a  
CC cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,  
CC c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and  
CC c-yes. The second oligonucleotide is specific for a nuclear oncogene or  
CC proto-oncogene selected from myc, jun, c-jun, c-fos, c-myc, B-myc,  
CC c-rel, c-vav, c-ski, c-spl, cyclin D1, PML/RAR alpha, AML1/MTG8,  
CC E2A/Prl and ALL-1/AF-4. The composition is used for treating cancer.  
CC The combination of antisense oligonucleotides has synergistically  
CC enhanced ability to inhibit growth of cancer cells.  
SQ Sequence 3623 BP; 843 A; 1090 C; 1086 G; 604 T;  
Query Match 5.1%; Score 79; DB 40; Length 3623;  
Best Local Similarity 56.7%; Pred. No. 1.08e-30;  
Matches 440; Conservative 0; Mismatches 331; Indels 5; Gaps 4;  
Db 833 ccccaactcagcaagtggagatggaacgcagcagcagcaccatcaccatgaagcacaagtgg 892  
QY 698 CCCCGAGGAGGAGCGTGTGGGAGCGGCACACTCCGAATTCGCCCTTGGGAGGAGCTGG 757  
Db 893 gcgggggccagtgcggggaggtgacagggcggtgtggaagaaatcacagctgcaggtgg 952  
QY 758 GTGAAGGCTACTTTGGGAGGTGTGGAGGCGCTGTGGCTGGCT-C--CTGCCCGTGG 814  
Db 953 cagtgaagacctgaaggaggacaccatcgaggtggaagagttcttgaagaagctgcag 1012  
QY 815 CGATCAAGGTTCATCAAGTCAGCAACATGAAGCTCACTGACCTGCCCAAGGA-GATCCAG 873  
Db 1013 tca-tgaagagatcaaaacctaactagtcagctcttgggtctgcaccggagag 1071  
QY 874 ACATGAAGGCGCTGGCGCAGCAGCGCGCTCATCCGGCTGCACGACGTGTCTCGGGCGGG 933  
Db 1072 ccccgcttctatcatcactgagttcatgacctacggaacctcctggaactactgagg 1131  
QY 934 GAGCCTGTGTACATAGTCAGGAACTCATCGCAAGGGGAACCTGAGGCTTCTCTGGGC 993  
Db 1132 gagtgaacccggcagaggtggaacgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1191  
QY 994 ACCCCGAGGGCGCGCCCTGCTGTGCGCCACTCTCTGGGCTTTGCTGCCAGGTGGCT 1053  
Db 1192 tcagccatggagtacatagagaagaaacttcacacagagatcttgtgccgaac 1251  
QY 1054 GAGGGATAGACTACCTTGGAGAGCAGCGCGTGTGTGACCGGACTTTGGCGCGCGGAAC 1113  
Db 1252 tgcctgtgagggagaacacctgtgtgaagtagctgtgtgtgtgtgtgtgtgtgtgtgt 1311  
QY 1114 GTGCTGTGACGACGCGCTTGGCTTGCAGAGGTGCTGACTTGGCCCTTGGCCGCTGCTC 1173  
Db 1312 acaggggacacctacacagcccagtcgtggagccaagtctcccatcaaatgactgcaccc 1371  
QY 1174 AAGGACGACATCTACTCCCGAGCAGCAGCTCCCAAGATCCCGGTCAAGTGCAGACGCGCT 1233  
Db 1372 gagagcctggcctacacaagttctccatcaagtccagctgtgtgtgtgtgtgtgtgtgtgt 1431  
QY 1234 GAGCGGCCAATATATGCTGTCTTCTCCAGAAAGTCAGACGCTGTGCTTCTGGCGGCTCTG 1293  
Db 1432 ctttgggaaattgctacctatgcatgtcccttacccttaccctgggaattgaccttccaggtg 1491  
QY 1294 CTCACGAGAGTTTTCACCTATGCGCAAGTGTCCCTATGAAGGGATGACCAACGAGGAGC 1353  
Db 1492 tatgagctgtagagaaggaactaccgcatgaagcgccaggaaggtgtcccgagaaaggtc 1551  
QY 1354 CTCACGACGATCATGGAGGGTACCAGGCTGCCGCGCGCGCGCTGCTTCCCGCGGAGGTC 1413

Db 1552 tatgaactcagcagcatgttggcagtggaatccctctgaccgcccctctcttgc 1607  
QY 1414 TACGTGCTCATGCTGGAGTGTGTGGAGGAGAGCCCCGAGGAGACGCCCTCTCTTTC 1469  
RESULT 10  
ID T61864 standard; DNA; 3780 BP.  
AC T61864;  
DT 06-JUN-1997 (first entry)  
DE c-abl gene.  
KW c-Abl gene; antisense RNA; cancer; therapy; DNA damage;  
KW radiotherapy; mitomycin C; cytosolic; tyrosine kinase; ss.  
OS Homo sapiens.  
PN W09708184-A1.  
PD 06-MAR-1997. U13922.  
PF 30-AUG-1996; U13922.  
PR 30-AUG-1995; US-520923.  
PA (ARCH-) ARCH DEV CORP.  
PA (DAND ) DANA FARBEN CANCER INST INC.  
PI Kharbanda S, Kufe DW, Weichselbaum RR;  
PI WPI; 97-179167/16.  
DR Antisense RNA molecule that inhibits expression of c-Abl gene  
PT Antisense RNA molecule that inhibits expression of c-Abl gene  
PT product - useful in killing malignant cells in cancer, optionally in  
PT combination with DNA damaging agent  
PS Example 9; Page 70-75; 114pp; English.  
CC The c-abl gene (T61864 and T61865) encodes a non-receptor tyrosine  
CC kinase that is activated DNA damaging agents. Antisense constructs  
CC directed to the c-abl gene can be used to selectively inhibit the  
CC expression c-Abl following exposure of cells to a DNA damaging  
CC agent such as ionising radiation or mitomycin C. When used in  
CC combination with DNA damaging agents, the antisense RNA provides a  
CC synergistic cell killing effect useful for killing malignant cells  
CC in cancer treatment.  
SQ Sequence 3780 BP; 853 A; 1136 C; 1166 G; 625 T;  
Query Match 5.1%; Score 79; DB 29; Length 3780;  
Best Local Similarity 56.7%; Pred. No. 1.08e-30;  
Matches 440; Conservative 0; Mismatches 331; Indels 5; Gaps 4;  
Db 1050 ccccaactcagcaagtggagatggaacgcagcagcagcaccatcaccatgaagcacaagtgg 1109  
QY 698 CCCCGAGGAGGAGCTGTGGGAGCGGCACACTCCGAATTCGCCCTTGGGAGGAGCTGG 757  
Db 1110 gcgggggccagtacggggaggtgtacagggcggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1169  
QY 758 GTGAAGGCTACTTTGGGAGGTGTGGAAAGCGCTGTGGCTGGCT-C--CTGCCCGTGG 814  
Db 1170 ccgtgaagacctgaagaggagacaccatggaggtggaagagttcttgaagaagctgcag 1229  
QY 815 CGATCAAGGTTCATCAAGTCAGCCCAACATGAAGCTCACTGACCTGCCCAAGGA-GATCCAG 873  
Db 1230 tca-tgaagagatcaaacacccctaacctagtcagctccttgggtgtgtgtgtgtgtgtgt 1286  
QY 874 ACATGGAAGGCTTGGCGCAGGAGCGGCTCATCCGGCTGACGAGTGTGCTCGGCGCGG 933  
Db 1289 ccccgcttctatcactgagttcatgacctacacgggaacctcctggaactacctgaggt 1348  
QY 934 GAGCCTGTGTACATAGTCACGGAACCTCATGCGCAAGGGGAACCTGACGCGCTTCTCTGGC 993  
Db 1349 gagtgaacccggcagggaggtgaacgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1408  
QY 994 ACCCCGAGGCGCGCGCGCTGCTGTGCGGCACTCTGCGGCTTGTGCTGCGAGGTGCT 1053  
Db 1409 tcagccatggagtagctacatagagaagaaacttcacacagatcttgtgtgtgtgtgtgtgt 1468  
QY 1054 GAGGCAATGACCTACCTGGAGGACAGCGGCTTGTGTACCGGGACTTGGCCGCGCGGAAC 1113  
Db 1469 tgcctgttaggggagaaccacttggtgaagtagctgtgtgtgtgtgtgtgtgtgtgtgtgt 1528  
QY 1114 GTGCTGTGGACGAGCGGCTGGCGCTGCAAGGTGCTGACTTCCGCTTGGCCCGGCTGCTC 1173  
Db 1529 acaggggacacctacacagcccagtcgtgtgagccaaagttcccccaataatgactgcaccc 1588

QY 1174 AAGGACGACATCTACTCCCGGAGCAGCAGCTCCAAAGATCCCGGTCAAGTGGACAGCGCCT 1233  
 Db 1589 gagaacctggcctacaaacaaagtcttcacatcaagtcacgaagtcgagcttggaatttgatg 1648  
 QY 1234 GAGCGCGCAATATATGCTGCTCTCTCCCAAGATCAGAGCTGTGCTCTCGCGCTCTG 1293  
 Db 1649 ctttgggaaatgtcacctatgtcattcccttaccctggaattgaccttccagggtg 1708  
 QY 1294 CTGACAGAGGTTTTCACCTATGCGCAGTCTCCATGAAGGATGACCAACACGAGAGC 1353  
 Db 1709 tatgagctgtgagaagaagactaccgcatgaagcgccgaagaggtgcccagagaaggtc 1768  
 QY 1354 CTGACAGCATCATGCGAGGGTACCGGCTGCCGCGCCGCTGCCCTGCCGCGGAGGTC 1413  
 Db 1769 tatgaactcatcgagcatgttggtgcaattggaatccctctgacggccctctttgc 1824  
 QY 1414 TAGCTGCTCATGCTGAGTGTGGAGGAGCAGCGCCCGGAGCAAGGCCCTCTTTCG 1469

## RESULT 11

ID Q30725 standard; DNA; 755 BP.  
 AC Q30725;  
 DT 22-MAR-1993 (first entry)  
 DE Potential tyrosine kinase receptor fragment RTK-3.  
 KW Tyrosine kinase receptor; ss.  
 OS Homo sapiens.  
 FH key Location/Qualifiers  
 FT cds 1..753  
 FT /\*tag= a  
 PN W09218149-A.  
 PD 29-OCT-1992.  
 PF 23-APR-1992; U03376.  
 PR 23-APR-1991; US-690199.  
 PR 26-JUL-1991; US-736559.  
 PA (REG- ) REGENERON PHARM INC.  
 PI Alarich TH, DiStephano P, Furth ME, Glass D, Masiakowski,  
 PI Maison-Pierre PC, Squinto SP, Stilt T, Yancopoulos GD;  
 DR WPI: 92-381778/46.  
 PT P-PSDB; R28605.  
 PT Assaying neurotrophin activity using cells that express trkB - also  
 PT for identifying agonists and antagonists, and new recombinant cells  
 PT and nucleic acid producing trkB, for diagnosis and treatment of  
 PT neurological disease  
 PS Disclosure: Fig 15; 156pp; English.  
 CC The sequence is that of potential tyrosine kinase receptor fragment  
 CC RTK-3, isolated from cDNA from human neuroblastoma cell line SY5Y. It  
 CC may be of use in the treatment of neurodegenerative disease/neurotrauma  
 CC including motor neuron disorders such as amyotrophic lateral sclerosis,  
 CC Werdnig-Hoffman disease, chronic proximal spinal muscular atrophy and  
 CC Guillain-Barre syndrome. It may also be of use in the treatment of  
 CC neurological diseases associated with diabetes, Parkinson's disease,  
 CC Alzheimer's disease and Huntington's chorea. See also Q30715-Q30728.  
 SQ Sequence 755 BP; 147 A; 257 C; 134 T;

Query Match 4.9%; Score 76; DB 5; Length 755;  
 Best Local Similarity 61.5%; Pred. No. 6.98e-29;  
 Matches 244; Conservative 0; Mismatches 150; Indels 3; Gaps 2;  
 Db 88 cagatcgccggcgagtgagtagtacctatccagccaccagctggttcacagacctggcc 147  
 QY 1045 CAGGTGGCTGAGGGATGAGTACTGTGGAGGAGCAGCGGCTGTGACCGGACTTGGCC 1104  
 Db 148 accccgaatgtctagtgtacgacaagctgaacgtgaagatctcagacttggcctcttc 207  
 QY 1105 GCCCGGACGCTGCTGTGGACGACGCGCTGGCTGCAAGGTTGCTGACTTGGCCCTGGCC 1164  
 Db 208 cgagaggtgtatgcccgcgattactacaagctgctggggaactcgtgctgctatccgc 267  
 QY 1165 CGGCTGCTCAAGGAGACATCTACTCCCCGA-GCAGCAG--CTCCAAGATCCCGGTCAG 1221  
 Db 268 tggatggccccagagccatcatgtacggaagttctccatcagactcagacatctggctc 327  
 QY 1222 TGGACAGCGCTGAGCGGCCCAATTATCGTCTTCTCCAGAAATCAGACGCTGGTCC 1341  
 Db 328 taagtggtgctgctgagaggtcttcacgtacgagctcagccctactcgcgggtattcc 387  
 QY 1282 TTCGGCGTCTCTGCACGAGGTTTTCACCTATGGCAGTGTGCCCTATGAAGGATGACC 1341  
 Db 388 aaccaggtgtggtgagatgatccggaacccgaggtgctgcttggccccgatgactgt 447  
 QY 1342 ACCACGAGACGCTGCAGCAGATCATGCGAGGGTACCGGGTGCCTGCTGCTGCTG 1401  
 Db 448 ccgcctgggtgtatgctccctcatgatcagtgctgga 484

Db 328 tacggtgtgctgctggtgaggtcttcaagctacgagcctcagccctactcgcgggtattcc 387  
 QY 1282 TTCGGCGTCTGCTGCACGAGGTTTTCACCTATGCGCAGTGTCCCTATGAAGGATGACC 1341  
 Db 388 aaccaggtgtggtgagatgatccggaacccgaggtgctgcttggccccgatgactgt 447  
 QY 1342 ACCACGAGACGCTGCAGCAGATCATGCGAGGGTACCGGCTGCGCGCGCTGCTGCTG 1401  
 Db 448 ccgcctgggtgtatgctccctcatgatcagtgctgga 484  
 QY 1402 CCGCGGAGGTTCTACGTGCTCATGCTGCTGAGTGTGGA 1438

## RESULT 12

ID V70224 standard; DNA; 755 BP.  
 AC V70224;  
 DT 11-FEB-1999 (first entry)  
 DE Rat orphan tyrosine kinase receptor RTK-3 encoding DNA.  
 KW Receptor tyrosine kinase; Ror-1; Ror-2; Ehk-1; Ehk-2; detection;  
 KW neurotrophin activity; trkB; proto-oncogene; tyrosine kinase receptor;  
 KW binding protein; BDNF; NT-3; diagnosis; ss.  
 OS Rattus sp.  
 PN US5843749-A.  
 PD 01-DEC-1998.  
 PF 06-JUN-1995; 469537.  
 PR 17-MAR-1995; US-406247.  
 PR 26-JUL-1991; US-736559.  
 PR 28-OCT-1993; US-144992.  
 PR 06-JUN-1995; US-469537.  
 PA (REG- ) REGENERON PHARM INC.  
 PI Maisonpierre PC, Masiakowski P, Yancopoulos GD;  
 DR WPI: 99-044584/04.  
 DR P-PSDB; W83166.  
 PT DNA encoding receptor tyrosine kinase proteins - and corresponding  
 PT proteins  
 PS Example: Fig 15; 194pp; English.  
 CC The present invention describes nucleic acid molecules for ror-1,  
 CC ror-2, ehk-1 and ehk-2. Also described are the corresponding proteins:  
 CC Ror-1; Ror-2; Ehk-1; and Ehk-2. The proteins are orphan receptor  
 CC tyrosine kinases. The present sequence encodes rat orphan tyrosine  
 CC kinase receptor Rtk-3 from the present invention.  
 SQ Sequence 755 BP; 147 A; 257 C; 134 T;

Query Match 4.9%; Score 76; DB 54; Length 755;  
 Best Local Similarity 61.5%; Pred. No. 6.98e-29;  
 Matches 244; Conservative 0; Mismatches 150; Indels 3; Gaps 2;  
 Db 88 cagatcgccggcgagtgagtagtacctatccagccaccagctggttcacagacctggcc 147  
 QY 1045 CAGGTGGCTGAGGGATGAGTACTGTGGAGGAGCAGCGGCTGTGACCGGACTTGGCC 1104  
 Db 148 accccgaatgtctagtgtacgacaagctgaacgtgaagatctcagacttggcctcttc 207  
 QY 1105 GCCCGGACGCTGCTGTGGAGGAGCAGCGGCTGGCTGCAAGGTTGCTGACTTGGCCCTGGCC 1164  
 Db 208 cgagaggtgtatgcccgcgattactacaagctgctggggaactcgtgctgctatccgc 267  
 QY 1165 CGGCTGCTCAAGGAGCAGATCTACTCCCCGA-GCAGCAG--CTCCAAGATCCCGGTCAG 1221  
 Db 268 tggatggccccagagccatcatgtacggaagttctccatcagactcagacatctggctc 327  
 QY 1222 TGGACAGCGCTGAGCGGCCCAATTATCGTCTTCTCCAGAAATCAGACGCTGGTCC 1281  
 Db 328 taagtggtgctgctgagaggtcttcacgtacgagctcagccctactcgcgggtattcc 387  
 QY 1282 TTCGGCGTCTCTGCACGAGGTTTTCACCTATGGCAGTGTGCCCTATGAAGGATGACC 1341  
 Db 388 aaccaggtgtggtgagatgatccggaacccgaggtgctgcttggccccgatgactgt 447  
 QY 1342 ACCACGAGACGCTGCAGCAGATCATGCGAGGGTACCGGGTGCCTGCTGCTGCTG 1401  
 Db 448 ccgcctgggtgtatgctccctcatgatcagtgctgga 484





\*\*\*\*\*  
M P S R E L F  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu May 20 13:27:11 1999; MacPar time 2933.07 Seconds  
Tabular output not generated. 1236.663 Million cell updates/sec

Title: >US-09-099-053-1  
Description: (1-1548) from US09099053.seq  
Perfect Score: 1548  
N.A. Sequence: 1 GCTCGCGGCTCCATCGGC  
Comp: CGAGCGCCGAGGATACCGG.....CCACGCTCGGCTCCAGC 1548

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
6:em\_est9 7:em\_gss1  
Database: genbank-est111  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_gss1 38:gb\_gss2  
39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 11.202; Variance 2.630; scale 4.260

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description Pred. No.
1	122	7.9	639 18	AA763708 vp06c03.r1 Soares mous 1.11e-151
2	110	7.1	547 21	AA940351 vz45g04.r1 Soares 2NbM 1.41e-131
3	98	6.3	336 12	AA354687 EST63065 Jurkat T-cell 1.02e-111
4	97	6.3	634 16	AA645158 vs36h03.r1 Stratagene 4.49e-110
5	93	6.0	553 20	AA869202 vq48h02.r1 Barstead bo 1.58e-103
6	92	5.9	577 11	AA306381 EST177341 Jurkat T-cell 6.77e-102
7	89	5.7	467 9	AA149096 z141g09.r1 Soares preg 5.12e-97
8	88	5.7	724 33	W226491 71811 Human retina cDN 2.14e-95
9	87	5.6	586 9	AA216991 mu04h04.r1 Soares mous 8.90e-94
10	83	5.4	519 13	AA420981 zt98e10.r1 Soares_test 2.50e-87

11	5.4	566 23	AA159486	vz75d05.r1 Soares mous 6.16e-89
12	81	5.2	AA356258	zr79f04.r1 Soares_NbM 4.02e-84
13	78	5.0	AA352185	EST60140 Activated T-c 2.48e-79
14	78	5.0	N32396	Yx73h12.r1 Soares mela 2.48e-79
15	75	4.8	AA098002	mm85b11.r1 Stratagene 1.40e-74
16	72	4.7	AA373274	qz48f07.x1 NCI_CGAP_K1 7.44e-70
17	73	4.7	T12334	AO23R Heart Homo sapie 2.00e-71
18	73	4.7	AA874687	v085a04.r1 Stratagene 2.00e-71
19	73	4.7	AA309943	EST180795 Jurkat T-cel 2.00e-71
20	70	4.5	AA035407	zk28h10.r1 Soares_preg 1.01e-66
21	69	4.5	AA794242	vu65a07.r1 Soares mous 3.64e-65
22	66	4.4	AA040695	zk49d09.r1 Soares_preg 1.31e-63
23	66	4.3	AA221421	mv22f12.r1 Barstead mo 1.63e-60
24	67	4.3	AA412971	ve94f03.r1 Knowles Sol 1.64e-62
25	65	4.2	AA571063	vk22b12.r1 Knowles Sol 5.68e-59
26	65	4.2	AA193845	qe71r11.x1 Soares_feta 5.68e-59
27	64	4.1	AA307109	v119a01.r1 Stratagene 1.95e-57
28	63	4.1	AA307109	EST178022 Colon carcin 6.65e-56
29	61	3.9	N98665	YV66d03.r1 Soares_mult 7.43e-53
30	59	3.8	AA203510	zx56a12.r1 Soares_feta 7.90e-50
31	58	3.7	T28388	EST40843 Human Skin Ho 2.53e-48
32	57	3.7	AA754459	97SN1787 Rice Immature 7.97e-47
33	57	3.7	W87315	zh64d04.r1 Soares_feta 7.97e-47
34	58	3.7	AI543823	LD39618.5prime LD Dros 2.53e-48
35	58	3.7	AA537193	vk45e12.r1 Soares mous 2.53e-48
36	58	3.7	AA266222	mz53a06.r1 Barstead mo 2.53e-48
37	57	3.7	AI005024	ou91a12.r1 NCI_CGAP_K1 7.97e-47
38	56	3.6	AA754459	97SN1787 Rice Immature 2.48e-45
39	56	3.6	F23146	SSC22G05 Porcine small 2.48e-45
40	52	3.4	M78193	EST01786 Subtracted Hi 1.99e-39
41	52	3.4	H75607	YU05h03.r1 Soares_feta 1.99e-39
42	53	3.4	AA446594	z9w0c01.r1 Soares_tot 6.80e-41
43	53	3.4	AA494631	qz17c04.x1 NCI_CGAP_CL 6.80e-41
44	51	3.3	H91029	YU89f07.r1 Soares_feta 5.70e-38
45	51	3.3	W62969	md88c09.r1 Soares mous 5.70e-38

ALIGNMENTS

RESULT 1  
LOCUS AA763708 639 bp mRNA EST 27-JAN-1998  
DEFINITION vp06c03.r1 Soares mouse mammary gland NbMNG Mus musculus CDNA clone IMAGE:1087812 5' similar to gb:J03023 Murine macrophage gene, encoding bmk (MOUSE);, mRNA sequence.  
ACCESSION AA763708  
NID g2813790  
VERSION AA763708.1 GI:2813790  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 639)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,F., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morrill,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797409.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.  
MGI:590172  
Seq primer: -28m13 rev2 ET from Amersham

SOURCE: house mouse.

SOURCE  
house mouse.

1266 C

T C A G A C G T C T G G T C C T T C G G C G T C C T G C T G C A C G

T C A G A C G T C T G G T C C T T C G G C G T C C T G C T G C A C G A G G T T T C A C C T A T G G C C A G T G T C 132

	Query Match	6.3%	Score 98	DB 12	Length 336	Best Local Similarity 65.7%	Pred. No. 1.02e-111	Mismatches 0	Mismatches 112	Indels 1	Gaps 1
Db	1	AAGATTCCAGACATTGGCCTTAGCAGCCTCATTTAGGAGCAACGAGTACACAGCCGGAG	60								
Qy	1141	AAGTGGCTGACTTCGGCTGGCCCGSCTGCTCAAGSAGACATCTACTCCCGGAGCAGC	1200								
Db	61	GGGGCCAAGTTTCCCATTAAGTGGACAGCGCCCAAGAGCCCAATTAACTACGGGACATTCACC	120								
Qy	1201	AGCTTCAAGATCCCGGTCTAGTGTGGACAGCGCTTAGCGCGCCCAATTTATCGTGTCTTCTCC	1260								

Qy	1261	CAGAAGTCAGACGCTCTGGTCTCTCGGCGTCTCTGCTGCACGAGGTTTTACCTATGGCCAG	1320
Db	181	ATCCCTTACCAGGGATGACCAACCCGGAGGTGATTCAGAACCTGGAGCGAGGCTACCGC	240
Qy	1321	TGTCCTTATGAAGGGATGACCAACCCAGACGCTGCAGAGATCATGCGAGGTCACCG	1380
Db	241	ATGCTGGCGCCCTGACAACTGTCAGAGGAGCTGTACCACTCAATTGAGGCTGTGCTGGAA	300
Qy	1381	CTGCGCGCGCGGTGCTGCCGGGGAGGTCTAGTGCTCATGCTGGA-GTGCTGGAG	1439
Db	301	GGAGCGCCCAAGAGCGCGCCACCTTTG	329
Qy	1440	GAGCAGCCCGGAGGAAGCGGCTCCTTTG	1468

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced q1:1407219.

EMAIL: moulees@wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:621613  
 Seq primer: -38ml3 rev1 ET from Amersham  
 High quality sequence stop: 321..  
 Location/Qualifiers  
 1..634  
 /organism="Mus musculus"  
 /note="Organ: blood; Vector: pBluescript SK-; Site:1:  
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;  
 Uni-rAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGC



```

3' -3' adaptor sequence: 5' CTCAGATTTTTTTTTTTTTTTTTT 3'
/bb_xref="taxon:10090"
/clone="IMAGE:1140341"
/clone_lib="Stratagene mouse Tcell 937311"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
150 a 171 c 169 g 144 t
BASE COUNT
ORIGIN

```

Query Match	6.38;	Score 97;	DB 16;	Length 634;
Best Local Similarity	64.28;	Pred. No. 4.49e-110;		
Matches	242;	Conservative 0;	Mismatches 133;	Indels 2; Gaps 2;
Db	5	CGCGCCCAACATCCTGTGTCTGACACGCTGAGCTGCAAGATTGCAGACTTTGGCTGCGC 64		
QY	1104	CGCCCGAAGCTGCTCGTGACACAGCGCCTGGCCTGCAAGTGGCTGACTTCGGCCTGGC 1163		
Db	65	CGGCTCATTTGAGCAAAATGAGTACAGCGCCCGGGAGGGGCCACAATTTCCCAATT-AGT 123		
QY	1164	CGGCTGCTCAAGGACGACATCTACTCC-CCGAGCAGCAGCTCCAAGATCCCGGTCAAGT 1222		
Db	124	GGACGACACAGAGGCCAATAACTATGGGACCTTCACCATCAAGTCAGAGGTGTGGTCT 183		
QY	1223	GGACAGCGCTGAGCGGGCCAAATTCTGTCTCTCCCAAGTCAGAGCTGTGGTCT 1282		
Db	184	TCGGGATCTTGTCTACAGATTGTCAACCCACGCGTGGAACTCCCTTACCAGGAATGACCA 243		
QY	1283	TCGGCGTCTGCTGCACGAGGTTTTCACTTATGCCAGTGTCCCTATGAAGGGATGACCA 1342		
Db	244	ACCCTGNAGTTCATTCAGAACTGGAGAGAGGCTACCGCATGGTGGACCTTGACAACTGT 303		
QY	1343	ACCACGAGACGCTGCAGCAGATCATCGAGGGTACCGCGTTCGCGGCCCGGCTGCGTGGC 1402		
Db	304	CGGAAGAGCTGTACCACTCATGATGCTGTGCTGGAAGGAGCGCCGCCAGAGGACCGGCCCA 363		
QY	1403	CGCGGAGGCTACGTTGCTCATCTGGAGTGTCTGGAGGACGACCCCGGAGGACGGCCCT 1462		
Db	364	CGTTTGAATACCTTCGG 380		
QY	1463	CTTTTGCACGCTCGG 1479		

RESULT	5
LOCUS	AA869202 553 bp mRNA EST 16-MAR-1998
DEFINITION	vq48n02.r1 Barstead bowel MPRRB9 Mus musculus cDNA clone IMAGE:1097523 5' similar to gb:M64608 Mouse lyn protein mRNA, complete cds (MOUSE);, mRNA sequence.
ACCESSION	AA869202
NID	G2964647
VERSION	AA869202.1 GI:2964647
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 553)
AUTHORS	Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilsson, R. and Waterston, R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2282426.

Contact: Marra M/Mouse EST Project  
WASHU-RHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:603755  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 505.

FEATURES	SOURCE
1. <b>Age</b>	1. <b>Age</b>
2. <b>Gender</b>	2. <b>Gender</b>
3. <b>Marital Status</b>	3. <b>Marital Status</b>
4. <b>Education</b>	4. <b>Education</b>
5. <b>Income</b>	5. <b>Income</b>
6. <b>Occupation</b>	6. <b>Occupation</b>
7. <b>Religion</b>	7. <b>Religion</b>
8. <b>Political Affiliation</b>	8. <b>Political Affiliation</b>
9. <b>Health Status</b>	9. <b>Health Status</b>
10. <b>Travel History</b>	10. <b>Travel History</b>
11. <b>Employment Status</b>	11. <b>Employment Status</b>
12. <b>Family Size</b>	12. <b>Family Size</b>
13. <b>Home Ownership</b>	13. <b>Home Ownership</b>
14. <b>Vehicle Ownership</b>	14. <b>Vehicle Ownership</b>
15. <b>Insurance Status</b>	15. <b>Insurance Status</b>
16. <b>Charitable Donations</b>	16. <b>Charitable Donations</b>
17. <b>Volunteer Work</b>	17. <b>Volunteer Work</b>
18. <b>Substance Use</b>	18. <b>Substance Use</b>
19. <b>Mental Health</b>	19. <b>Mental Health</b>
20. <b>Physical Activity</b>	20. <b>Physical Activity</b>
21. <b>Dietary Habits</b>	21. <b>Dietary Habits</b>
22. <b>Smoking Status</b>	22. <b>Smoking Status</b>
23. <b>Alcohol Consumption</b>	23. <b>Alcohol Consumption</b>
24. <b>Stress Levels</b>	24. <b>Stress Levels</b>
25. <b>Social Media Usage</b>	25. <b>Social Media Usage</b>
26. <b>Travel Frequency</b>	26. <b>Travel Frequency</b>
27. <b>Home Security</b>	27. <b>Home Security</b>
28. <b>Vehicle Maintenance</b>	28. <b>Vehicle Maintenance</b>
29. <b>Insurance Premiums</b>	29. <b>Insurance Premiums</b>
30. <b>Charitable Contributions</b>	30. <b>Charitable Contributions</b>
31. <b>Volunteer Hours</b>	31. <b>Volunteer Hours</b>
32. <b>Substance Use Frequency</b>	32. <b>Substance Use Frequency</b>
33. <b>Mental Health Screenings</b>	33. <b>Mental Health Screenings</b>
34. <b>Physical Activity Levels</b>	34. <b>Physical Activity Levels</b>
35. <b>Dietary Intake</b>	35. <b>Dietary Intake</b>
36. <b>Smoking Status</b>	36. <b>Smoking Status</b>
37. <b>Alcohol Consumption</b>	37. <b>Alcohol Consumption</b>
38. <b>Stress Levels</b>	38. <b>Stress Levels</b>
39. <b>Social Media Usage</b>	39. <b>Social Media Usage</b>
40. <b>Travel Frequency</b>	40. <b>Travel Frequency</b>
41. <b>Home Security</b>	41. <b>Home Security</b>
42. <b>Vehicle Maintenance</b>	42. <b>Vehicle Maintenance</b>
43. <b>Insurance Premiums</b>	43. <b>Insurance Premiums</b>
44. <b>Charitable Contributions</b>	44. <b>Charitable Contributions</b>
45. <b>Volunteer Hours</b>	45. <b>Volunteer Hours</b>
46. <b>Substance Use Frequency</b>	46. <b>Substance Use Frequency</b>
47. <b>Mental Health Screenings</b>	47. <b>Mental Health Screenings</b>
48. <b>Physical Activity Levels</b>	48. <b>Physical Activity Levels</b>
49. <b>Dietary Intake</b>	49. <b>Dietary Intake</b>
50. <b>Smoking Status</b>	50. <b>Smoking Status</b>
51. <b>Alcohol Consumption</b>	51. <b>Alcohol Consumption</b>
52. <b>Stress Levels</b>	52. <b>Stress Levels</b>
53. <b>Social Media Usage</b>	53. <b>Social Media Usage</b>
54. <b>Travel Frequency</b>	54. <b>Travel Frequency</b>
55. <b>Home Security</b>	55. <b>Home Security</b>
56. <b>Vehicle Maintenance</b>	56. <b>Vehicle Maintenance</b>
57. <b>Insurance Premiums</b>	57. <b>Insurance Premiums</b>
58. <b>Charitable Contributions</b>	58. <b>Charitable Contributions</b>
59. <b>Volunteer Hours</b>	59. <b>Volunteer Hours</b>
60. <b>Substance Use Frequency</b>	60. <b>Substance Use Frequency</b>
61. <b>Mental Health Screenings</b>	61. <b>Mental Health Screenings</b>
62. <b>Physical Activity Levels</b>	62. <b>Physical Activity Levels</b>
63. <b>Dietary Intake</b>	63. <b>Dietary Intake</b>
64. <b>Smoking Status</b>	64. <b>Smoking Status</b>
65. <b>Alcohol Consumption</b>	65. <b>Alcohol Consumption</b>
66. <b>Stress Levels</b>	66. <b>Stress Levels</b>
67. <b>Social Media Usage</b>	67. <b>Social Media Usage</b>
68. <b>Travel Frequency</b>	68. <b>Travel Frequency</b>
69. <b>Home Security</b>	69. <b>Home Security</b>
70. <b>Vehicle Maintenance</b>	70. <b>Vehicle Maintenance</b>
71. <b>Insurance Premiums</b>	71. <b>Insurance Premiums</b>
72. <b>Charitable Contributions</b>	72. <b>Charitable Contributions</b>
73. <b>Volunteer Hours</b>	73. <b>Volunteer Hours</b>
74. <b>Substance Use Frequency</b>	74. <b>Substance Use Frequency</b>
75. <b>Mental Health Screenings</b>	75. <b>Mental Health Screenings</b>
76. <b>Physical Activity Levels</b>	76. <b>Physical Activity Levels</b>
77. <b>Dietary Intake</b>	77. <b>Dietary Intake</b>
78. <b>Smoking Status</b>	78. <b>Smoking Status</b>
79. <b>Alcohol Consumption</b>	79. <b>Alcohol Consumption</b>
80. <b>Stress Levels</b>	80. <b>Stress Levels</b>
81. <b>Social Media Usage</b>	81. <b>Social Media Usage</b>
82. <b>Travel Frequency</b>	82. <b>Travel Frequency</b>
83. <b>Home Security</b>	83. <b>Home Security</b>
84. <b>Vehicle Maintenance</b>	84. <b>Vehicle Maintenance</b>
85. <b>Insurance Premiums</b>	85. <b>Insurance Premiums</b>
86. <b>Charitable Contributions</b>	86. <b>Charitable Contributions</b>
87. <b>Volunteer Hours</b>	87. <b>Volunteer Hours</b>
88. <b>Substance Use Frequency</b>	88. <b>Substance Use Frequency</b>
89. <b>Mental Health Screenings</b>	89. <b>Mental Health Screenings</b>
90. <b>Physical Activity Levels</b>	90. <b>Physical Activity Levels</b>
91. <b>Dietary Intake</b>	91. <b>Dietary Intake</b>
92. <b>Smoking Status</b>	92. <b>Smoking Status</b>
93. <b>Alcohol Consumption</b>	93. <b>Alcohol Consumption</b>
94. <b>Stress Levels</b>	94. <b>Stress Levels</b>
95. <b>Social Media Usage</b>	95. <b>Social Media Usage</b>
96. <b>Travel Frequency</b>	96. <b>Travel Frequency</b>
97. <b>Home Security</b>	97. <b>Home Security</b>
98. <b>Vehicle Maintenance</b>	98. <b>Vehicle Maintenance</b>
99. <b>Insurance Premiums</b>	99. <b>Insurance Premiums</b>
100. <b>Charitable Contributions</b>	100. <b>Charitable Contributions</b>

```

i: 1253
/organism="Mus musculus"
/strain="FVB/N"
/note="Vector: p7T3D-Pac (
polylinker; Site_1: EcoRI;
was primed with a Not I - o
TGTTCAGCAATCTGAAGTGGGAGCGGCC
3'); double-stranded CDNA
[AAATCGGATCCTTG], digested
Not I and Eco RI sites of
Source irradiated bowel har
irradiation (1400 Gys). Li
Barstead.
/db_xref="taxon:10090"
/clone="IMAGE:1097523"
/clone_lib="Barstead_bowel_1
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
142 a 142 g 122 t
BASE COUNT

```

BASE COUNT	142 a	144 c	145 g	122 t
BASE COUNT	142 a	144 c	145 g	122 t

Query Match	5.08;	Score 93;	DB 20;	Length 553;	
Best Local Similarity	60.3%;	Pred. No. 1.58e-103;			
Matches	272;	Conservative	0;	Mismatches 179;	Indels 0; Gaps 0;

Db	96	TGAAGACCTTGCAACATGACAAAGCTAGTGGCGGTGTACGCTGTGGTGTACCAAGGAGGAGC	155
QY	878	TGAAGGCCCTGCGCAGCAGCGGCTCATCGCGTGTACGCAGTGTGCTCGGGGGGAGC	937
Db	156	CCATCTACATCATACCGAGTTTCATGGCTAAGGTAGTTTGCCTGGATTTCCTCAAGAAGTG	215
QY	938	CTGTGTACATAGTCACGGAACATCATGCGAAGGGAACCTGCAGGCGCTTCTGGGCACCC	997
Db	216	ATGAAGGTGGCAAGGTGCTCTGCCCAAGTCTCATTTGACTTCTCGGCCACAGATTCGCAAG	275
QY	998	CCGAGGCCGGGGCCCTGCGTCTGCCGCCACTCTCTGGGCTTTGCTGCGAGGTGGCTGAGG	1057
Db	276	GCATGGCGTACATCGAGCGGGAAGAACTACATCCACCGTGATCTCGAGCTGTCAACATCC	335
QY	1058	GCATGAGCTACCTGGAGGACGCGGTTGTGCACCGGACTTGGCCGCCCGGAACGTGC	1117
Db	336	TGCTCTCTGAGTCACATCATGTGCAAGATTCACAGACTTTGGCCTCGCGAGAGTCAATCGAAG	395
QY	1118	TCGTGGACGACGGCCCTGGCCTGCAAGGTGGCTGACTTCGGCCCTGGCCCGGCTGCTCAAGG	1177
Db	396	ATAACGAGTACACAGCAAGGAAGGTGCGAAGTTCCCTATCAAGTGGACAGCTCCAGAGG	455
QY	1178	ACGACATCTACTCCCGAGCAGCAGCTCCAAGATCCCGGTCAAGTGGACAGCGCGCTGAGG	1237
Db	456	CCATCAACTTCGGCTGCTTCACTATCAAAATCTGACGTGTGGTCCCTTCGGAAATTCCTCGT	515
QY	1238	CGGCCAATTATCGTGTCTCTCCAGGAAGTCACAGCTCTGGTCTCTTCGGGCGTCTCGCTGC	1297
Db	516	ATGAGATTGTCACCTATGGGAAGATCCCTA	546
QY	1298	ACGAGGTTTTACCTATGGCCAGTGTCCCTA	1328

RESULT	6				
LOCUS	AA306381	577 bp	mRNA	EST	18-APR-1997
DEFINITION	EST177341	Jurkat T-cells VI	Homo sapiens cDNA 5'	end similar to	
				similar to tyrosine kinase 19kDa	
ACCESSION	AA306381				
NID	91958708				
VERSION	AA306381.1	GI:1958708			
KEYWORDS	EST.				

6	AA306381	577 bp	mrna	EST	18-APR-1997
LOCUS	EST177341	Jurkat T-Cells	VT Homo sapiens	DDNA 5'	end similar to
DEFINITION		similar to tyrosine kinase lck, mRNA sequence.			
ACCESSION	AA306381				
NID	91958708				
VERSION	AA306381.1	GI:1958708			
KEYWORDS	EST.				

```

SOURCE
ORGANISM      human.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;
Eutheria: Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 577)
AUTHORS
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai.C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., HungJun,J., Li,H., Meissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
TITLE
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL
Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE
96026280
COMMENT
On Sep 21, 1992 this sequence version replaced gi:276421.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699036
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tldb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1..577
/organism="Homo sapiens"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
/db_xref="ATCC (inhost):160506"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells V1"
/cell_type="T-lymphocyte"
144 a 163 c 149 g 120 t 1 others
BASE COUNT
ORIGIN
Query Match 5.9%; Score 92; DB 11; Length 577;
Best Local Similarity 60.7%; Pred. No. 6.77e-102;
Matches 328; Conservative 0; Mismatches 207; Indels 5; Gaps 5;

Db 36 TGAAGACGTGCAACACACCGCGCTGGTTCGGCTCTACGCTGTG-G-TCACCCAGG-AGC 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 878 TGAAGGCGCTGCGGCAGCAGCGCTCATCGCGCTGCACGAGTGTGTCGGCGGGAGC 937
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93 CCATCTACATCATCATGATACATGAGATGGAGTCTAGTGGATTCTCAAGACCC 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 938 CTGTGTACATAGTCACGGAACACTATCGCAAGGGGAAACCTGCGAGCGCTTCTGGGCACCC 997
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 CTTCAAGCATCAAGTTGACCATCAACAACTCTCTGGACATGGGACGCCAAATTCGAGAAG 212
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 998 CCGAGGCGCGGCGCTCGCTCTGCCCGCACTCTCTGGCGTTGCTGCGCAGGTGCTGAGG 1057
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 213 GCATGGCATTCATTGAAGACGGAATATATTCATCGTGTACCTTCGGGCTGCCAACATTC 272
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1058 GCATGAGCTACCTGGAGGACGAGCGCGCTTGTGCACCGGGACTTGGCCGCCCGGAACCTGC 1117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 273 TGTGTCTGTGACACCTGAGCTGGAAGATTGACAGACTTGGCTAGACAGCGCTCATTTAGG 332
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

FEATURES
  source
    seq primer: 66cgaagaaagccttgaagaaattt.
    Location/Qualifiers
      1..724
        /organism="Homo sapiens"
        /note="Organ: eye; Vector: lambda gt10; Site_1: EcoRI; Site_2: EcoRI. The library used for sequencing was a sublibrary derived from a human retina cDNA library. Inserts from retina cDNA library DNA were isolated, cleaved with tsp5091, size selected, and cloned into lambda gt10. Individual plaques were arrayed and used as templates for PCR amplification and these PCR products were used for sequencing."
        /db_xref="taxon:9606"
        /clone_lib="Human retina cDNA Tsp5091-cleaved sublibrary"
        /sex="mixed (males and females)"
        /tissue_type="retina"
        /dev_stage="adult"
        /lab_host="E. coli strain K802"
  source
    seq primer: -28ml3 rev2 ET from Amersham
    Location/Qualifiers
      1..586
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

```

TITLE The WasU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)

COMMENT On Jan 19, 1998 this sequence version replaced qi:2152293.

Contact: Marra M/Mouse EST Project  
WASHU-HHMI Mouse EST Project  
Washington University School  
of Medicine  
4444 Forest Park Parkway, Box 9501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:691841

Trace considered overall poor quality  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 1.

## FEATURES

source

```

1. 566
/organism="Mus musculus"
/strain="C57BL/6J"
/note="organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(3n) primer [5',
TGTACCAATCTGAAGTGGCGCGCGCGAATGGTTTTTTTTTTTTTTTTTTT
T3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonalde."

```

BOUATTO.  
/db xref="taxon.10090"

```
/cd_xref = caxon:10030  
/clone="IMAGE:1332297"
```

```
/clone_lib="Soares mouse mammary gland NbMMG"
```

/sex="male"

```
/tissue_type="mammary gland"
```

```
/dev_stage=
```

```
lab_host="DH10B"
```

141 a 153 c 150 g 122 t

ORIGIN

Query Match	5.48:	Score 84:	DB 23:	Length 566:
-------------	-------	-----------	--------	-------------

Best Local Similarity 62.1%; Pred. No. 6.16e-89;

Matches	216;	Conservative	0;	Mismatches	132;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	------	--------	----	------	----

Db 78 CACCGAGACTTGAGGGCAGCCACACATCCTGGTGGGGGAATACCTAATATGCAAGATCGCT 137

Qy 1090 CACCGGACTTGGCCGCCCGGACGTCGTGGACGACGGCCTGGCCTGCAAGGTGGCT 1149

138 GACTTCGGGCTGGGCACGGCTCCTAAGTGACAAATGAGTATAACCCCCACACAGGAACCAAG 197

100

QY 1150 GACTTCGGCCTGGCCCGGCTGCTCAAGGACGACATCTACTCCCGAGCAGCAGCTCCAAG 1209

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

DB 198 T T C C C C A T C A A G T G G A C A G C C C C A G A G C C G C C C T C T T T G G C A G A T T C A C T G T C A A A T C A 257

QV 1210 ATCCCGGTCAAGTGGACAGCGCCTGAGGCGGCCAATTATCGTGTCTTCTCCAGAAGTCA 1269

[illegible]

Db 258 GACGTGTGTCCTTTGGGATTCTGCTCACTGAACATGATCACCAAGGCAGAGTTCCTTAC 317

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

RESULT 12

LOCUS	AA256258	574 bp	mrna	EST	06-AUG-1997
DEFINITION	zr79f04.r1	Soares_NIHMPU_S1	Homo sapiens cDNA clone IMAGE:681919		
	similar to	gb:M14676	PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN		

```

Db 392 GTGCTGGACAGGTGGAGCGAGCTACAGGATGCCCTGCCCGCAGGACTGCCCATCTCT 451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1351 ACGCTGCACGAGTATCGGAGGTACCGCTGCCCGCCGCTGCCCTGCCCGGGAG 1410

Db 452 CTGCGATGACATGATCCACTGCTGGAAAAAGGACCCCTGGAGAGCGCCCATCTTTG 509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1411 GTCTAGTCTCATGCTGGAGTGTGGAGGAGCAGCGCCGCGAGGAACGGCCCTCTTTG 1468

RESULT 13
LOCUS AA352185 334 bp mRNA EST 21-APR-1997
DEFINITION EST60140 Activated T-cells XX Homo sapiens cDNA 5' end similar to
        tyrosine kinase lck, mRNA sequence.
ACCESSION AA352185
NID 92004505
VERSION AA352185.1 GI:2004505
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiiridae; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
        Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
        White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
        Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
        Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
        Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
        Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrilgrino,J.M.,
        Moreno-Palances,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
        Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
        Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
        Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
        Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
        He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
        Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
        Raymond,L., Wei,X.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
        Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
        Fraser,C.M. and Venter,J.C.
        Initial assessment of human gene diversity and expression patterns
        based upon 83 million nucleotides of cDNA sequence
        Nature 377 (6547 Suppl), 3-174 (1995)
JOURNAL 96026280
MEDLINE
COMMENT On Nov 29, 1993 this sequence version replaced gi:430407.
        Other_ESTs: THC170446
        Contact: Kerlavage, AR
        Bioinformatics
        The Institute for Genomic Research
        9712 Medical Center Drive, Rockville, MD 20850 USA
        Tel: 3018699056
        Fax: 3018699423
        Email: arkerlav@tigr.org
        For clone availability, additional sequence and expression
        information related to this EST, please check the TIGR Human Gene
        Index (http://www.tigr.org/tadb/hgi/hgi.html)
        Seq primer: M13 Reverse.
        Location/Qualifiers
            1..334
            /organism="Homo sapiens"
            /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
            XhoI"
            /db_xref="ATCC (inhost):152375"
            /db_xref="taxon:9606"
            /clone_lib="Activated T-cells XX"
            /cell_type="T-lymphocyte"
            /dev_stage="adult"
            70 a 98 c 96 g 69 t 1 others
BASE COUNT 70 a 98 c 96 g 69 t 1 others
ORIGIN

Query Match 5.0%; Score 78; DB 12; Length 334;
Best Local Similarity 67.4%; Pred. No. 2.46e-79;
Matches 151; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

```

Db 18 TTCCACATCAAGTCAGATGTGTGCTTTTGGGATCCTGCTGACGGAATTTTCACCCAC 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1255 TTCCACAGAGTCAGACGCTGTGCTCTGGCGCTCTGCTGCACGAGTTTTCACCTAT 1314

Db 78 GGCCGCATCCCTTACCACGGGATGACCAACCCGAGAGTGATTTCAGAACCTTGAGCGAGGC 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1315 GGCCAGTGTCCCTATGAAGGATGATCACCACACGAGAGCGTGCAGCAGATCATCGAGGG 1374

Db 138 TACCGATGTGTGGCCCTGACAACTGTCCAGAGGAGCTGTACCAACTCATGAGCGCTGTGC 197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1375 TACCGGCTGCCCGCGCGGCTGTCTGCCCGCGGAGGTACGTGCTCATGCTGGAGTGC 1434

Db 198 TGAAGGAGGCGCCAGAGGACCGGCCACCTTTGACTACCTGCG 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1435 TGGAGGAGCAGCCCGGAGGAACGCCCTCTTTGCCACGCTGCG 1478

RESULT 14
LOCUS N32396 401 bp mRNA EST 10-JAN-1996
DEFINITION Yx73h12.r1 Soares melanocyte 2NBHM Homo sapiens cDNA clone
        IMAGE:267431 5' similar to gb:MI4676 PROTO-ONCOGENE
        TYROSINE-PROTEIN KINASE FYN (HUMAN);, mRNA sequence.
ACCESSION N32396
NID g1152795
VERSION N32396.1 GI:1152795
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiiridae; Homo.
REFERENCE 1 (bases 1 to 401)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
        Hollan,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
        Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
        Trevasaki,E., Waterston,R., Williamson,A., Wohldmann,P. and
        Wilson,R.
        The WashU-Merck EST Project
        Unpublished (1995)
        On Apr 14, 1993 this sequence version replaced gi:785619.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 257
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1009 Std Error: 0.00
Seq primer: T7
High quality sequence stop: 257.
        Location/Qualifiers
            1..401
            /organism="Homo sapiens"
            /note="Vector: pT73D (Pharmacia) with a modified
            polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5',
            TGTTCACAACTGAGTGGAGCGCGCGAGTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was size selected, ligated to Eco RI
            adapters (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of a modified pT73 vector
            (Pharmacia). Library constructed by Beato Soares and
            M.Fatima Bonaldo. RNA from normal foreskin melanocytes
            (FS374) was kindly provided by Dr. Anthony P. Albino."
            /db_xref="GDB:3877073"
            /db_xref="taxon:9606"
            /map="17q21"
            /clone="IMAGE:267431"
            /clone_lib="Soares melanocyte 2NBHM"
            /sex="Male"
            /tissue_type="melanocyte"
FEATURES
        source

```

BASE COUNT 106 a 99 c 109 g 83 t 4 others  
ORIGIN

Query Match 5.0%; Score 78; DB 32; Length 401;  
Best Local Similarity 62.1%; Pred. No. 2,46e-79;  
Matches 223; Conservative 0; Mismatches 134; Indels 2; Gaps 2;  
Db 8 AACATTCTAGTGGGAATGGAAGTCAATATGCAAGATGCTGACTTGGATGGCCGGATG 67  
Qy 1111 AACGTGCTGGTGGAGCGGCTGGCCCTCAAGTGGGTGACTTCGGCTGGCCGGGTG 1170  
Db 68 ATAGAAGCAATGAGTACACAGCAGCAAGGTCAAGTCCCATCAAGTGGACGGCC 127  
Qy 1171 CTCAGGACGACATCTACTCCCCGAGCAGCAGCTCCAGATCCCCGTCAAGTGGACAGCG 1230  
Db 128 CCCGAGCAGCCCTGTACGGAGGTTCAATCAAGTCTGAGTGTGTCTTTTGGAAATC 187  
Qy 1231 CCTGAGGGGCAATATATCGTGTCTTCTCCAGAGTCAAGAGTCTGTCTTCGGCGTC 1290  
Db 188 TTACTCAGAGCTGGTCAACAAAGAGAGTGCATACCCAGGATGAACAAACGGGAG 247  
Qy 1291 CTGTGTCAGCAGAGTTTTCACCTATGCGCAGTGTCCCTATGAAGGATGACCAACACAG 1350  
Db 248 TGCCTGGAGAGTGGAGCGAGGCTACAGGATGCCCTGCCCG-CAGACTGCCCTATCT 306  
Qy 1351 ACGCTGCAGCAGATCATCGAGGGTACCGGTGCCCGCGCGGTGCTGCCCGCGGAG 1410  
Db 307 CTGCATGAGCTCATGATCCATCTGTGGAAAAAGNCCCTGGAAGACGCCCCACTTTTG 365  
Qy 1411 GTCTACGTGCTCATGCTGGAGTGTGGAGGAGCAGCCCG-AGGAACGGCCCTCCTTTG 1468

RESULT 15  
LOCUS AA098002 562 bp mRNA EST 15-FEB-1997  
DEFINITION m85b11.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone  
IMAGE:550845 5' similar to gb:M12056 Mouse rearranged lck gene  
encoding lymphocyte-specific protein (MOUSE);, mRNA sequence.  
ACCESSION AA098002  
NID g1643555  
VERSION AA098002.1 GI:1643555  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 562)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393736.

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:331637  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 333.  
FEATURES  
SOURCE 1..562  
Location/Qualifiers  
/organism="Mus musculus"  
/note="Organ: blood; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:

Oligo dT, M30 CD4+ cells. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGACACGAG  
3' -3' adaptor sequence: 5' CTCAGTTTTTTTTTTTTTTT 3',  
/db\_xref="taxon:10090"  
/clone="IMAGE:550845"  
/clone\_lib="Stratagene mouse Tcell 937311"  
/tissue\_type="Tcell"  
/dev\_stage="M30 CD4+ cells"  
/lab\_host="SOLR (kanamycin resistant)"  
BASE COUNT 130 a 156 c 143 g 132 t 1 others  
ORIGIN

Query Match 4.8%; Score 75; DB 36; Length 562;  
Best Local Similarity 65.3%; Pred. No. 1.40e-74;  
Matches 171; Conservative 0; Mismatches 90; Indels 1; Gaps 1;  
Db 31 AAGTGGACAGCACCAAGACCCATTAACTATGGACCTTCAACCATCAAGTCAGACGTGTGG 90  
Qy 1219 AAGTGGACAGCCCTGAGCGGCCCAATATCTGTCTTCCAGAGTCAGACGTCTGG 1278  
Db 91 TCCTTTCGGGATCTTCTTATCAGAGATTGTACCCACGGTCAATCCCTTACCCAGGAAT 150  
Qy 1279 TCCTTTCGGGCTCCTGCTGCAC-GAGGTTTTCACCTATGSCCAGTGTCCCTATGAAGGAT 1337  
Db 151 GACCAACCCCTGAAGTCATTCAGAACTGGAGAGAGGCTACCCGATGGTGAGACCTGACAA 210  
Qy 1338 GACCAACCCAGAGAGCTGCAGCAGATCATGCGAGGGTACCGGCTGCCCGCGGCTGC 1397  
Db 211 CTGTCCGGAAGAGCTTACCACCTCATGATGCTGTCTGGAAGAGCGCCCGAGAGCCG 270  
Qy 1398 CTGCCCCGGGAGGTCTAGTGTCTCATGCTGTGAGTGTCTGGAGAGCAGCCCGAGGAGC 1457  
Db 271 GCCCACGCTTTGACTACCTTCGG 292  
Qy 1458 GCCCTCCTTTGCCACGCTGCGG 1479

Search completed: Thu May 20 14:58:13 1999  
Job time : 5462 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:13:59 1999; MasPar time 28.99 Seconds

Tabular output not generated. 357.975 Million cell-updates/sec

Title: >US-09-099-053-2

Description: (1-488) from US09095053.pep (1 of 6)

Perfect Score: 3671

Sequence: 1 MEPLRRRLAFLSFWDKIW .....ERPSPATLREKLHAIRCHP 488

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a:geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 36.486; Variance 158.060; scale 0.231

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3671	100.0	488	39	Human SAD	0.00e+00
2	1355	36.9	451	12	Breast tumor kinase,	7.03e-117
3	1332	36.3	505	8	PTK gene LptK-2 prod.	1.25e-114
4	1332	36.3	505	14	Protein tyrosine-kinase	1.25e-114
5	1326	36.1	533	8	Chicken pp60 c-src pr	4.84e-114
6	1315	35.8	536	8	Human pp60 c-src prot	5.77e-113
7	1189	32.4	417	3	(Beta-galactosidase N	1.17e-100
8	1114	30.3	1146	3	Abelson Related Gene,	2.39e-93
9	1072	29.2	1182	3	Abelson Related Gene,	2.92e-89
10	962	26.2	259	6	Sequence of pp60(c-sr	1.40e-78
11	926	25.2	620	17	ITK tyrosine kinase,	4.29e-75
12	904	24.6	659	17	BTK tyrosine kinase,	5.76e-73
13	900	24.5	441	17	Drosophila Src28C tyr	1.40e-72
14	894	24.4	466	13	N-terminal truncated	5.34e-72
15	894	24.4	507	13	Cytoplasmic tyrosine k	5.34e-72
16	893	24.3	659	20	Mouse haematopoietic-	6.67e-72

17	890	24.2	507	15	R84181	Megakaryocyte kinase	1.30e-71
18	882	24.0	630	17	R94536	TEC tyrosine kinase.	7.71e-71
19	873	23.8	675	15	R84182	Megakaryocyte kinase	5.71e-70
20	873	23.8	675	17	R94533	BMX tyrosine kinase.	1.70e-66
21	837	22.8	528	34	W64454	Human matk protein.	8.48e-59
22	757	20.6	246	13	R71131	Tyrosine kinase domai	1.20e-57
23	745	20.3	928	19	R97853	Rat REK7 eph-related	2.91e-57
24	741	20.2	877	35	W71628	Mouse Bsk receptor-l1	5.64e-57
25	738	20.1	991	15	R85090	EPH-like receptor pro	1.70e-56
26	738	20.1	1005	38	W83147	Rat receptor tyrosine	1.70e-56
27	733	20.0	710	13	R75714	Eph-related PTK Cdk7,	1.70e-56
28	733	20.0	722	13	R75705	Eph-related PTK Cdk7,	1.70e-56
29	733	20.0	744	13	R75713	Eph-related PTK Cdk7+	3.30e-56
30	730	19.9	983	6	R31466	HEK polypeptide.	2.41e-55
31	721	19.6	380	8	R44512	Elk PTK.	2.41e-55
32	721	19.6	984	8	R44513	elk.	4.66e-55
33	717	19.5	998	19	W03421	Mouse developmental k	5.81e-55
34	718	19.4	951	13	R75704	EPH-like receptor pro	2.18e-54
35	711	19.4	951	13	R75704	Eph-related CEK6	1.75e-54
36	712	19.4	983	13	R75711	Eph-related PTK Cdk4.	2.18e-54
37	711	19.4	994	24	W26366	Mouse Nuk tyrosine ki	4.23e-54
38	708	19.3	994	15	R87018	Receptor tyrosine kin	2.72e-54
39	710	19.3	995	13	R75712	Eph-related PTK Cdk5.	2.72e-54
40	710	19.3	1011	13	R75709	Eph-related PTK Cdk5+	6.57e-54
41	706	19.2	849	13	R75706	Eph-related PTK Cdk8.	8.19e-54
42	705	19.2	970	15	R85089	EPH-like receptor pro	6.57e-54
43	706	19.2	986	14	R85936	Protein tyrosine-kin	6.57e-54
44	706	19.2	986	15	R85091	EPH-like receptor pro	5.95e-53
45	696	19.0	973	13	R75707	Eph-related PTK Cdk9.	

## ALIGNMENTS

RESULT 1

ID W89248 standard; Protein; 488 AA.

AC W89248; 1999 (first entry)

DE Human SAD.

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease.  
OS Homo sapiens.  
PN W09849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGEN) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plozman GD;  
DR WPI; 99-009434/01.  
DR N-PSDB: V81743.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Claim 2; Page 154-155; 193pp; English.

CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC present sequence represents human SAD. The above proteins, other than  
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
CC substances that modulate their activity (i.e. agonists and antagonists,  
CC including NBP) in vivo or in vitro. These substances are used to treat  
CC or prevent diseases associated with abnormal signal transduction  
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
CC serine/threonine kinase) are used to promote neuronal survival,  
CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
CC diseases. Nucleic acid fragments of the polynucleotides encoding the



CC proteins can be used as probes to identify and clone related sequences;  
 CC to detect protein-encoded RNA; to generate transgenic animals and in  
 CC gene therapy (optionally after mutation). Ab are used to determine the  
 CC proteins.  
 SQ Sequence 488 AA;

Query Match 100.0%; Score 3671; DB 39; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
 Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mepflrrlatifffdwkiwaggepdhgtgslpntdpvtlpaecspfpqflaly 60  
 QY 1 MEPFLRRRLAFLSFDWKIWPAGGPDHGTGSLDPNTDPTVTLPAECSPFPQFLALY 60

Db 61 dftarcggelsvrrdricalaeeggyifarrlsgqpsaglvpthvakaspetsldgwp 120  
 QY 61 DFTARC GGELS VRRDRICALAE EGGYIFARRLSGQPSAGLV PTHVAKAS PETSLDGPW 120

Db 121 yfsgvrtgaqqlilspnpegafliirpsesslggyslvraqakvchyrvsmaadgsly 180  
 QY 121 YFSGVRSRTQAQQLILSPNPEGAF LIIRPSESSLGGYSLVRAQAKVCHYRVSM AADGS LY 180

Db 181 lqkgrlfpqgleelltykanwklignllqpcmpkqprqdwvwrphsefalkrklgey 240  
 QY 181 LQKGR LFPGLEELLTYKANWKLI GNLLQPCMPKQPRQDWVWRPHSEFAL KRKLGEY 240

Db 241 fgevweglwslpvaikvksanmkltdlakeiqltqlgrlherlirlhavcsggepyvi 300  
 QY 241 FGEVWEGWLW SLPVAIKVKSANMKLTD LAKEIQLTQLGR LHERLIRLHAVCSGGE PYVI 300

Db 301 vtelmrknqlcagltgpegralrllpqlgfacqvaegmsyleeqrvhrdlaarnvlvdd 360  
 QY 301 VTELMRK NQLCAGLTGPEGRALRLLPQLG FACQVAEGMSYLEEQRVHRDLAARNV LVD D 360

Db 361 glackvadfglarllkddispsssskipkwtapeaanyrvfsgksdwsfgvllhev 420  
 QY 361 GLACKVADFG LARLLKDDISPSSSSKIP KWTAPEAANYRVFSGKSDWSFGVLLHE V 420

Db 421 tygpcyegmtnhetlqimrgyrlprpaacpaevyvimlcwrrsspeerspfatlrekl 480  
 QY 421 TYGPCYEGMTNHETLQOIMRGYRLPRPAACPAEVYVIMLCWRRSSPEERSPFATLREKL 480

Db 481 haihrchp 488  
 QY 481 HAIHRCHP 488

RESULT 2  
 ID R63088 standard; Protein; 451 AA.  
 AC R63088; 1995 (first entry)  
 DT 12-AUG-1995  
 DE Breast tumor kinase, brk.  
 KW Breast tumor kinase; brk; protein-tyrosine-kinase; PTK;  
 KW Breast cancer; metastasis; prognosis; diagnosis; T-47D.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 15..68  
 FT /note= "similar to PTK SN3 domain"  
 FT domain 78..169  
 FT /note= "similar to PTK SH2 domain"  
 FT binding\_site 198..203  
 FT /note= "putative ATP binding domain"  
 FT domain 312..317  
 FT /note= "corresponds to the strong indicator  
 FT sequence of ptk specificity in  
 FT subdomain VI"  
 FT modified\_site 342  
 FT /note= "putative autophosphorylation site"  
 FT misc\_difference 447  
 FT /note= "potential regulatory tyrosine equivalent  
 FT to Tyr-527 of c-src"  
 PN W09502057-A.  
 PD 19-JAN-1995.

PF 08-JUL-1994; G01479.  
 PR 09-JUL-1993; GB-014233.  
 PR 11-MAR-1994; GB-004817.  
 PA (CANC-) CANCER RES INST.  
 PA (WELL) WELLCOME FOUND LTD.  
 PI Barker KT, Crompton MR, Gusterson BA, Kamalati T;  
 PI Mitchell PJ, Page MJ, Spence P;  
 DR WPI; 95-066901/09.  
 DR N-PSDB; Q81189.  
 PT Novel protein tyrosine kinase and its DNA - isolated from human  
 PT breast tumour, useful for diagnosis and prognosis of cancerous  
 PT tissue  
 PS Disclosure: Page 29-31; 52pp; English.  
 CC A brk (breast tumor kinase) cDNA fragment was used to screen a  
 CC cDNA library prepared from human breast tumor cell line T-47D. 2  
 CC cDNAs were isolated. The slightly longer clone lambda-t2 (Q81189)  
 CC encoded brk (863088), which was identified as a novel putative  
 CC non-receptor kinase of use as a prognostic/diagnostic of breast  
 CC tumor metastasis.  
 SQ Sequence 451 AA;

Query Match 36.9%; Score 1355; DB 12; Length 451;  
 Best Local Similarity 46.5%; Pred. No. 7.03e-117;  
 Matches 202; Conservative 73; Mismatches 151; Indels 8; Gaps 8;

Db 13 yvglwdfkrtdeelsfragdvfhvarkeeqwattlldeagavaggyvphnylaeret 72  
 QY 56 FLALYDFTARC GGELSVRRDRICALAE EGGYIFARRLSGQPSAGLV P-ITHVAKAS PET 114

Db 73 vesepwffgicirseaarrllqaegnagatgflirvsekspsadyvlsvrdtqavthykiwrr 132  
 QY 115 LSDQPMYFSGVSRTOAQQLLSPNPEGAF LIIRPSESSLGGYSLVRAQAKVCHYRVSM A 174

Db 133 aggrlhlneavsfslpelvnyhraq-slshgrlraapcrkhepeplphwdwverpreef 191  
 QY 175 ADGSLYLQGR LFPGLEELLTYKANWKLI GN-PLLOPCMPQKA-P-RQ-DWVERHSEF 230

Db 192 tlcrklsgygfegfeglwkdrrvqvaikvisrdnllhqqmlqseigamkkltrhkhaly 251  
 QY 231 ALGRKLGEYFGEVWEGWLW SLPVAIKVKSANMKLTD-LAKEIQLTQLGR LHERLIRLH 289

Db 252 avsvsgdpvyiitelmakgslllellrdsdeklvpvseellidwqvaegmcylesqvihr 311  
 QY 290 AVCSGGEPYIIVTELMRKGNLQAF LGTPEGRALRLLPQLG FACQVAEGMSYLEEQRVHR 349

Db 312 dlaarnilvgentlckvgdflarllkddvy-lshdhnipykwtapealsrghystksdv 370  
 QY 350 DLAAARNVLVDDGLACKVADFG LARLLKDDIYSPSSSSKIP KWTAPEAANYRVFSKSDV 409

Db 371 wsfgillhemfsgvpypgmsnheafirvdagymrmpcpleppsvhklmltcwcrdpeq 430  
 QY 410 WSFGVLLHEVFTYGC PYEGMTNHETLQOIMRGYRLPRPAACPAEVYVIMLCWRRSSPEE 469

Db 431 rpcfkalerlsssf 444  
 QY 470 RPSFATLREKLHAI 483

RESULT 3  
 ID R41941 standard; Protein; 505 AA.  
 AC R41941;  
 DT 10-MAR-1994 (first entry)  
 DE PTK gene LptK-2 prod.  
 KW PTK; protein tyrosine kinase; catalytic domain; c-kit; megakaryocyte;  
 KW lymphocyte; amplification; primer; polymerase chain reaction; PCR.  
 OS Homo sapiens.  
 PN W09315201-A.  
 PD 05-AUG-1993.  
 PF 22-JAN-1993; U00586.  
 PR 22-JAN-1992; US-826935.  
 PA (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.  
 PI Avraham H, Cowley S, Groopman J, Scadden D;  
 DR WPI; 93-320330/40.

[illegible]









Db 222 iceakvmtklqhnplvqlvgvctkhrpiiviteymkhgallnlyrrhextlignmglld 281  
 QY 271 AKEIOTLGLRHERLIRLHAVCGSGEPVIVITELMRKGNLQALFLGTPEGRAL-RLPPLLG 329  
 Db 282 mciqyskgtlylryhirdlaarnclygsenvkvadglarylddgytssggtkfp 341  
 QY 330 FACQVAEGMSYLEEQRVVRDARLVVDDGLACKVADFGRLARLKKDDIYSPSSSKIP 389  
 Db 342 ikwappevlnytrfasksdvaygvilmwefctgkmpygrlknvtevervqgillekpk 401  
 QY 390 VKWTAPEAANYRVFSQKSDWSEFVLLHEVFTYGCQPYEGMTNHETLQOIMRGYRLPR 449  
 Db 402 scakeiydmklcwsghgeepafvrlmqdl 432  
 QY 450 ACPAEVYVLMLECWRSPEERPSFATLREKL 480

## RESULT 14

ID R71132 standard; Protein: 466 AA.  
 AC R71132;  
 DT 27-OCT-1995 (first entry)  
 DE N-terminal truncated cytoplasmic tyrosine kinase.  
 KW cytoplasmic; tyrosine kinase; blood; cell differentiation;  
 KW screening; anticancer agent; SH3; src homology domain.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 7..70  
 FT domain 81..155 /note="SH3 domain"  
 FT domain 192..438 /note="SH2 domain"  
 FT domain 192..438 /note="tyrosine kinase domain"  
 PN W09506113-A.  
 PD 02-MAR-1995.  
 PF 25-AUG-1994; J01411.  
 PR 25-AUG-1993; JP-210403.  
 PR 29-MAR-1994; JP-058553.  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 PI Sakano S;  
 PI WPI: 95-106842/14.  
 DR N-PSDB; Q84888.  
 PT Cytoplasmic tyrosine kinase and antibody recognising it - for  
 PT screening chemical substances for tyrosine kinase inhibitory or  
 PT activating activity for use as cancer therapy  
 PS Claim 1; Page 40-42; 58pp; English.  
 CC A cytoplasmic tyrosine kinase which has enhanced expression in  
 CC connection with blood cell differentiation has been isolated from the  
 CC human UT-7 blood cell line. This sequence comprises an N-terminal  
 CC truncated fragment of the enzyme (residues 42-507 of R71133). The DNA  
 CC sequences and antibodies raised against the enzyme, are useful for  
 CC screening agents for inhibiting or activating activity on the tyrosine  
 CC kinase, for use as anticancer agents.  
 SQ Sequence 466 AA;

Query Match 24.4%; Score 894; DB 13; Length 466;  
 Best Local Similarity 39.8%; Pred. No. 5,34e-72;  
 Matches 169; Conservative 78; Mismatches 155; Indels 23; Gaps 16;

Db 24 gelafrkdvtilleacecnksyrvkhhtsgggellaaalrealsadpklslmpwf 83  
 QY 68 GELSVRRGDRLCAL-EGGG-YIFARRLSGQPSA-GLVPITHVAKASPET-LSQDPWF 122  
 Db 84 gkisgeavqql-qpped-glflvresarhpdgylcvsgfgrdvihrv-lhrdghtld 140  
 QY 123 SGVSTQAOQLLSPNPEGAFLIRPSESLGGYSLVRAQAKVCHYRVSMADGSLYLQ 182  
 Db 141 eavffcnlmdmvehyskdkgaictklvrpkrhgtksaeelaragwllnqlhtlgaqi 200  
 QY 183 KGRLPFGEELTYTYKANWKLQNPLQ-CMP-QKAPQDV---WERPHSEFALGRKL 236  
 Db 201 gegefavlgqeylqg-kvavknkic-dvtaqafldetavmtkmqhenlvrlglilh-q 257

QY 237 GEGYFGEVNEGLWGLSLPVAIKVKSANKMLTDLAKEIOTLGLRHERLIRLHAVCSGGE 296  
 Db 258 glyivmehvskgnlvnfltr-gralvntaqllqfslhaegmeyleskklvhrdlaarn 316  
 QY 297 PVIVITELMRKGNLQALFLGTPEGRAL-RLPPLLGACQVAEGMSYLEEQRVVRDLAARN 355  
 Db 317 ilvseodlvakvsdfglakaerkg1----dsrllpvtwapealkhgtksdwsfgvl 372  
 QY 356 VLVDDGLACKVADFGRLARLKKDDIYSPSSSKIPVKTAPFAAANYRVFSQKSDWSEFVL 415  
 Db 373 lvevfysgrapykmslkevsaeavekgyrmepegcpgvhlmscweaeaparrpfrk 432  
 QY 416 LHEVFTYGCQPYEGMTNHETLQOIMRGYRLPRPAACPAEVYVLMLECWRSPEERPSFAT 475  
 Db 433 laekl 437  
 QY 476 LREKL 480

## RESULT 15

ID R71133 standard; Protein: 507 AA.  
 AC R71133;  
 DT 27-OCT-1995 (first entry)  
 DE Cytoplasmic tyrosine kinase.  
 KW cytoplasmic; tyrosine kinase; blood; cell differentiation;  
 KW screening; anticancer agent; SH3; src homology domain.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 48..111  
 FT domain 122..196 /note="SH3 domain"  
 FT domain 233..478 /note="SH2 domain"  
 FT domain 233..478 /note="tyrosine kinase domain"  
 PN W09506113-A.  
 PD 02-MAR-1995.  
 PF 25-AUG-1994; J01411.  
 PR 25-AUG-1993; JP-210403.  
 PR 29-MAR-1994; JP-058553.  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 PI Sakano S;  
 PI WPI: 95-106842/14.  
 DR N-PSDB; Q84888.  
 PT Cytoplasmic tyrosine kinase and antibody recognising it - for  
 PT screening chemical substances for tyrosine kinase inhibitory or  
 PT activating activity for use as cancer therapy  
 PS Claim 1; Page 42-44; 58pp; English.  
 CC A cytoplasmic tyrosine kinase which has enhanced expression in  
 CC connection with blood cell differentiation has been isolated from the  
 CC human UT-7 blood cell line. The DNA sequences and antibodies raised  
 CC against the enzyme, are useful for screening agents for inhibiting or  
 CC activating activity on the tyrosine kinase, for use as anticancer agents.  
 SQ Sequence 507 AA;

Query Match 24.4%; Score 894; DB 13; Length 507;  
 Best Local Similarity 39.8%; Pred. No. 5,34e-72;  
 Matches 169; Conservative 78; Mismatches 155; Indels 23; Gaps 16;

Db 65 gelafrkdvtilleacecnksyrvkhhtsgggellaaalrealsadpklslmpwf 124  
 QY 68 GELSVRRGDRLCAL-EGGG-YIFARRLSGQPSA-GLVPITHVAKASPET-LSQDPWF 122  
 Db 125 gkisgeavqql-qpped-glflvresarhpdgylcvsgfgrdvihrv-lhrdghtld 181  
 QY 123 SGVSTQAOQLLSPNPEGAFLIRPSESLGGYSLVRAQAKVCHYRVSMADGSLYLQ 182  
 Db 182 eavffcnlmdmvehyskdkgaictklvrpkrhgtksaeelaragwllnqlhtlgaqi 241  
 QY 183 KGRLPFGEELTYTYKANWKLQNPLQ-CMP-QKAPQDV---WERPHSEFALGRKL 236  
 Db 242 gegefavlgqeylqg-kvavknkic-dvtaqafldetavmtkmqhenlvrlglilh-q 298  
 QY 237 GEGYFGEVNEGLWGLSLPVAIKVKSANKMLTDLAKEIOTLGLRHERLIRLHAVCSGGE 296





\*\*\*\*\*

protein - protein database search, using Smith-Waterman algorithm

```
Run on: Thu May 20 12:18:07 1999; MasPar time 28.76 Seconds
        635.602 Million cell-updates/sec
Tabular output not generated.
```

```
>US-09-098-053-2
Description: (1-488) from US0909053.pep (1 of 6)
Perfect Score: 3671
Sequence: 1 MEPEFLRRRLALISFFWDKIW.....ERPSEATLREKLAHTRCHP 488
```

Scoring table: PAM 150  
Gap II

Searched: 116695 seqs, 37453910 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

```
Database:
pir60
1:pir1 2:pir2 3:pir3 4:pir4
```

Statistics: Mean 49.350: Variance 96.105: scale 0.513

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2938	80.0	496	2	A56040	protein-tyrosine kina
2	2932	79.9	496	2	I56222	srn - mouse
3	1442	39.3	506	2	S24553	protein-tyrosine kina
4	1391	37.8	505	2	S24550	protein-tyrosine kina
5	1388	37.9	507	2	A39939	protein-tyrosine kina
6	1355	36.9	451	2	S49016	protein-tyrosine kina
7	1341	36.5	505	1	TVHUHC	protein-tyrosine kina
8	1340	36.5	509	2	I48845	p56-lck - mouse
9	1339	36.5	509	2	A23639	protein-tyrosine kina
10	1335	36.4	512	2	I49552	protein-tyrosine kina
11	1336	36.4	529	1	TVHUFR	protein-tyrosine kina
12	1332	36.3	505	1	I38396	protein-tyrosine kina
13	1333	36.3	509	1	OKHULG	protein-tyrosine kina
14	1333	36.3	532	2	B34104	protein-tyrosine kina
15	1329	36.2	532	2	A34104	protein-tyrosine kina
16	1324	36.1	509	1	TVHAST	protein-tyrosine kina
17	1326	36.1	533	1	TVCHS	protein-tyrosine kina
18	1326	36.1	568	1	TVFVSV	protein-tyrosine kina
19	1321	36.0	503	2	JO1321	protein-tyrosine kina
20	1321	36.0	557	1	TVFVSV	protein-tyrosine kina
21	1321	36.0	587	1	TVFVPR	protein-tyrosine kina
22	1317	35.9	503	1	TVMSGH	protein-tyrosine kina
23	1306	35.6	526	2	S20808	protein-tyrosine kina



```
Db 394 GAKFPIKWTAPAEAMYSRFTIKSDWSFGVLVYEVITYGREPYGPMTNAAVLQIQOSYR 453
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 385 SSKIPVKTAPAEANRYRVFSQKSDWSFGVLLHEVFTYGCPEYEGMNHETLQOIMRGYR 444
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 454 MPRPMGCPERLYAIMDCWDREDPASRPFTFL 485
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 445 LPRPAACPAEVVLMLECWRRSSPEERPSFATL 476
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
ENTRY #24550 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) 1 - freshwater sponge
      (Spongilla lacustris)
ALTERNATE_NAMES src-type tyrosine kinase 1
ORGANISM #formal_name Spongilla lacustris
DATE 07-May-1993 #sequence_revision 07-May-1993 #text_change
      08-Sep-1997
ACCESSIONS S24550
REFERENCE #molecule_type mRNA
#authors Raulf, F.
#submission submitted to the EMBL Data Library, September 1991
#accession S24550
#residues 1-505 #label RAU
#cross-references EMBL:X61601; NID:g10149; PID:g10150
GENETICS
#gene srk1
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
      homology; SH2 homology; SH3 homology
KEYWORDS ATP; phosphotransferase; tyrosine-specific protein kinase
FEATURE 61-111 #domain SH3 homology #label SH3\
      122-214 #domain SH2 homology #label SH2\
      238-496 #domain protein kinase homology #label KIN\
      246-254 #region protein kinase ATP-binding motif\
      268 #active_site Lys #status predicted
SUMMARY #length 505 #molecular_weight 57693 #checksum 3389
Query Match 37.9%; Score 1391; DB 2; Length 505;
Best Local Similarity 43.6%; Pred. No. 1.19e-260;
Matches 197; Conservative 95; Mismatches 147; Indels 13; Gaps 8;

Db 39 EPKPKPPQVPPAQDVKYP-IYGVKDYDSRTDDDLSEKKGDLMIYIISTDGDWFWFARSK 97
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 35 DPNTDPVPTLPFAPCSPFPOLFALYDFTARCGLSVRRGDRLEALE-EGGGYIFARRL 93
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 98 DTAGKEGYIPSNVA-EY-KSLDAEEFLGKIKRVEAEKMLNQSNQVGSFLIRDSSTTP 155
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 94 SGQPSAGLVPIITHVAKASPTLSDQPWYFSGVSRTOAQQLLSPNPEGAFILRPSESL 153
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156 GDFSLSVKDDRVHRVRRLEDSLFVTRRSTFQIHLVDHYKIETDGLCCCKLLYPCL 215
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 154 GYSLSVRAQAKVCHYRVSMADGSLYLQKGRLPFGLELLTYKANKWKLQNLQPCM 213
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 216 QAEKPTAGLLRQANWEIEKTOIKLLRIGAQGFGEWGLNGTTSVAVKTLKPGTM 275
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 214 ----PQKAP--RQ--DWVERPHSEFALGRKLGEYFGVEWGLMGLSLPVAIKSANM 265
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 276 SVBEFLQEAESIMKRLRPKLIQIYAVCTKEEPIYIVTLMKYGSLLEYLRGDS-VLKIE 334
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 266 KLTDLAKEIQTGLGRHERLIRLHAVCSGGPEYIVTLMKRLQALFLGTPGGRALRLP 325
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 335 QLVDDVAQAAGSGLYLBQONVIRDLARNILVGEHGCTKADFGLARVIDEIEYEAHTG 394
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 326 PLGFCACQVAGHSYLEEQVRVHRDLARNVLVDDGLACKVADFGLARLLKADDDIYSRSS 385
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 395 AKFPIKWTAPAEAMYNRTIKSDWSFGVLLVYEVITYGREPYGPMTNPEVLEKIQOYRM 454
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 386 SKIPVKTAPAEANRYRVFSQKSDWSFGVLLHEVFTYGCPEYEGMNHETLQOIMRGYRL 445
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 455 PCFANCPKQFHDIMDCWDREDPASRPFTFLQ 486
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 446 PRPAACPAEVVLMLECWRRSSPEERPSFATLR 477

RESULT 5
ENTRY #A39939 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) tk1 - chicken
      kinase-related transforming protein (tk1); T-cell surface
      antigen associated protein tk1
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 06-Mar-1992 #sequence_revision 05-May-1995 #text_change
      07-Nov-1997
ACCESSIONS A42126; A39939
REFERENCE #molecule_type mRNA
#authors Chow, L.M.; Ratcliffe, M.J.; Veillette, A.
#journal Mol. Cell. Biol. (1992) 12:1226-1233
#title tk1 is the avian homolog of the mammalian lck tyrosine
      protein kinase gene.
#cross-references MUID:92186854
#accession A42126
#residues 1-88 #label CHO
#cross-references GB:M85043
#experimental_source thymus, spleen
#note sequence extracted from NCBI backbone (NCBIN:88831,
      NCBIP:88833)
REFERENCE A39939
#authors Streibhardt, K.; Mullins, J.I.; Bruck, C.; Ruebsamen-Waigmann,
      H.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:8778-8782
#title Additional member of the protein-tyrosine kinase family: the
      src- and lck-related protooncogene c-tkl.
#cross-references MUID:88097370
#accession A39939
#status preliminary
#molecule_type mRNA
#residues 52-507 #label STR
#cross-references GB:J03579; NID:g212712; PID:g212713
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
      homology; SH2 homology; SH3 homology
KEYWORDS ATP; lipoprotein; myristylation; phosphoprotein;
      phosphotransferase; surface antigen; tyrosine-specific
      protein kinase
FEATURE 66-114 #domain SH3 homology #label SH3\
      125-222 #domain SH2 homology #label SH2\
      241-499 #domain protein kinase homology #label KIN\
      249-257 #region protein kinase ATP-binding motif\
      2 #modified_site myristylated amino end (Gly) (in mature
      form) #status predicted
SUMMARY #length 507 #molecular_weight 58011 #checksum 4536
Query Match 37.8%; Score 1388; DB 2; Length 507;
Best Local Similarity 45.3%; Pred. No. 5.25e-260;
Matches 202; Conservative 93; Mismatches 136; Indels 15; Gaps 10;

Db 55 PCSPLQDKLVVLYDYEPTHDGLKQGEKLRVLVEESGEWWRQAQLTGQ-EGLIPIHF 113
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 48 PCSFPFPQ-LFLALYDFTARCGLSVRRGDRLEALEGGGYIFARRLSQPSAGLVPITH 106
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 114 VAMVN--SLEPEPFFKNLSRKNARLLASGNTHGSLFIRESETSKGYSLSVRDFDQ 171
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 107 VAKASPTLSQPWYFSGVSRTOAQQLLSPNPEGAFILRPSESSLGYSLSVRA---- 162
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 172 QGEVTKVHKIRNMONGGYII-SPRVTFSSLHELVEYSSSDGLCTRLGKPCRTQPKP 230
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 163 QA-KVCHFRVMAADGSLYLQKGR-L-FPGLELLTYKANKWKLQNLQPCMPQKA--P 218
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 WQDEWEVPRESLKVEKLGAQGFGEWVMGYNHGTVAIKNLKOGSMSPSAFLAEANLM 290
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 219 R-QDWERPHSEFALGRKLGEYFGVEWGLWGLSLPVAIKVKSANKMLTDLAKEIQT 277
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 KNLQHPRLVRLYAVTK-EPIYIITEYMEKSLVDLFTKSETGKLSINKLDMAAQIAEG 349
      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



[illegible]

regions are present in t cells.

```
##cross-references MUID:88142832
#accession I57629
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-11 #label RES
#cross-references GB:M18098; NID:g198766; PID:g198767
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
KEYWORDS ATP; autophosphorylation; kinase-related transforming
protein; phosphoprotein; phosphotransferase;
tyrosine-specific protein kinase
FEATURE
68-116 #domain SH3 homology #label SH3\
127-224 #domain SH2 homology #label SH2\
243-501 #domain protein kinase homology #label KIN\
251-259 #region protein kinase ATP-binding motif
SUMMARY #length 509 #molecular-weight 57952 #checksum 3251
Query Match 36.5%; Score 1339; DB 2; Length 509;
Best Local Similarity 44.4%; Pred. No. 1.85e-249;
Matches 198; Conservative 91; Mismatches 142; Indels 15; Gaps 10;
Db 57 PASPLQDNLVIALHSYPSHDGDLGFEKGEQLRILEQSGEWKKAQSLTGTQ-EGTIPFNF 115
Qy 48 PCSFPQQLFTALYDFTARCGLSVRRGDRLCALCEEGGYIFARRLSQPSAGLVPIITH 106
Db 116 VAKAN--SLEPEPFPKRLSRKDAERQQLIAPGNTHGSLFIRESESTAGSFLSRVDFDON 173
Qy 107 VAKASPEITSDQPWYFSGVSRTOAQLLLSPNPGAFILRSESSLGYSLSVRA-- 162
Db 174 QGVVYKHKIRNLNDNGYI--SPRTTFGLDLVRYHYNASGLCTKTSRQCOTQKPKP 232
Qy 163 QAKVC-HYRVSMADGSLYLQKGR-L-FPGLELLYYKANWKLQNLPLQLQPCMPQKA--P 218
Db 233 WNEDEWEVPRTLKLVRLGAGQCFQGVWYNGYHGTAKVSKLQSGSPVFLAEANLM 292
Qy 219 R-QDWERPHESEFALGRKLGEYFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIQT 277
Db 293 KQLQPRLVRLYAVVTO-EPIYIIIEYMENGLVDFLTPSGIKLVNKLDDMAQIABG 351
Qy 278 KGLRHLRLHLCVSGGEPYIVTELKRNQNLQAFGLTPEGRALRLPLGFCQVABG 337
Db 352 MAFTIEQNYIHRDLRAANILVSDLSCKIADFGLARLIEDNYTTAREGAKFPKWTAPEA 411
Qy 338 MSYLEEQRVVHRDLAARNVLVDGDLACKVADFGLARLLKDDIYSPSSSKIPVKWTAPEA 397
Db 412 INYGTFTIKSDWSFGIILLTEIVTHGRIPYGMNTNPEVIONLGRYRVPRPNCPEELYH 471
Qy 398 ANYRVFSOKSDWSFGVLLHEVFTYGQCPYEGMTNHTLQOIMRGYRLPRPAACPAEVV 457
Db 472 LMMLCKWRPERDPRTFYLRSLVDDF 497
Qy 458 LMLECWRSPPERPSPFATLRKLHAI 483
RESULT 10
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) bsk/iyk - mouse
ALTERNATE_NAMES intestinal tyrosine kinase
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
09-Apr-1998
ACCESSIONS I49552; I48608
REFERENCE I49552;
#authors Oberg-Welsh, C.; Welsh, M.
#journal Gene (1995) 152:239-242
#title Cloning of BSK, a murine FRK homologue with a specific
pattern of tissue distribution.
#cross-references MUID:95137395
#accession I49552
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
```

```
##residues 1-512 #label RES
##cross-references GB:L36132; NID:g556287; PID:g777773
REFERENCE I48608
#authors Thureson, M.; Albrecht, D.; Zurcher, G.; Andres, A.C.;
Ziemiecki, A.
#journal Biochem. Biophys. Res. Commun. (1995) 209:582-589
#title iYk, a novel intracellular protein tyrosine kinase
differentially expressed in the mouse mammary gland and
intestine.
#cross-references MUID:95251656
#accession I48608
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-153, 'T', 155-236, 'H', 238-512 #label RE2
#cross-references EMBL:Z48757; NID:g736263; PID:g736264
GENETICS BSK
#gene #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
CLASSIFICATION ATP; intestine; phosphotransferase
KEYWORDS #domain SH3 homology #label SH3\
FEATURE 56-112 #domain SH2 homology #label SH2\
123-215 #domain protein kinase homology #label KIN\
239-501 #region protein kinase ATP-binding motif
247-255 #length 512 #molecular-weight 58891 #checksum 3691
SUMMARY
Query Match 36.4%; Score 1335; DB 2; Length 512;
Best Local Similarity 43.2%; Pred. No. 1.35e-248;
Matches 198; Conservative 102; Mismatches 140; Indels 18; Gaps 16;
Db 38 PBAPRSQEPERSHGQYFVALFDYQARTAEEDLSFRAGDKLQVLDTSHEGWLAHLE-KKG 96
Qy 40 PVTPLPAEPCSPFPQLFLALYDFTARCGLSVRRGDRLCALCEEG-GGYIFARRLSQPS 98
Db 97 TGLGQOOLYIISYVAEDRSLOAEFWFPGAKRADAEKQLLYSENOTGAFILRESESQK 156
Qy 99 AGLVP-I-THV-AK-ASPE-TLSDQWPYFSGVSRTOAQLLLSPNPGAFILRSESSL 153
Db 157 GDFSLVLDGEGVVYKHVIRRLDEGGFFLTRKVFSTLNFVNYYTTTSDGLCVKLEPKCL 216
Qy 154 GYISLSVRAQAKVCHVRYVSMADGSLYLQKGRFLPGLELLYYKANWK-L-I--QNP-L 208
Db 217 KIQVTPFDLSYKTADQWEIDRNSIQLLKRLSGQGEVWGLWNNNTTPVAVTKLPKGM 276
Qy 209 -LQPCMPQKAP-R-QDWERPHESEFALGRKLGEYFGEVWGLWGLSLPVAIKVKSANM 265
Db 277 DPNDFLRAEQIMKSLRHPKLIQLYAVCTLEDPIYIITELMRHGSLOEYLDNDGSGSKIHF 336
Qy 266 KLTDLAKEIQTILGRHERLIRLHVCSGGEPPYIVTELKRNQNLQAFGLTPEGRALRLP 325
Db 337 QQVDMAAQVAGSMAYLESQNYIHRDLAARNVLGHEINIKVADFGLARVFKVNDIEDIYES 396
Qy 326 PLUGFACQVAGESYLEDQGVVHRDLAARNVLDDGLACKVADFGLARLK-D--DIYSP 382
Db 397 KHEIKLPVKWTAPETAIKTKSIKSDWSFGIILYIIITYGKMPYSGMTGAQVIQMLSQN 456
Qy 383 SSSSKIPVKWTAPETAIKTKSIKSDWSFGVLLHEVFTYGQCPYEGMTNHTLQOIMRG 442
Db 457 YRLPQSPNCQPOFYSLMLECWNVEPKQRTFETLHWKL 494
Qy 443 YRLPRPAACPAEVVLMLECWRSPPERPSPFATLRKL 480
RESULT 11
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) fgr - human
ALTERNATE_NAMES kinase-related transforming protein (fgr)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Dec-1988 #sequence_revision 30-Sep-1989 #text_change
22-May-1998
ACCESSIONS A27676; A28353; A24842; A45930; S24306
REFERENCE A27676
```







243-501 #domain protein kinase homology #label KIN\

251-259 #region protein kinase ATP-binding motif\

2 #modified\_site myristylated amino end (Gly) (in mature form) #status predicted\

3,5 #binding\_site palmitate (Cys) (covalent) #status predicted\

273 #active\_site Lys #status predicted\

394 #binding\_site phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted\

505 #binding\_site phosphate (Tyr) (covalent) #status predicted\

SUMMARY #length 509 #molecular-weight 58000 #checksum 1143

Query Match 36.3%; Score 1333; DB 1; Length 509;

Best Local Similarity 44.48; Pred. No. 3,62e-248;

Matches 198; Conservative 91; Mismatches 142; Indels 15; Gaps 10;

Db 57 PASPLQDNLVIALHSYBPSHDGIDFGKGEQLRIELEGSGEWKQAQSLTTGO-EGFIPFNF 115

QY 48 PCSFPFQ-LFLALYDFATRCGGELSVRRGRLCALEBGGVIFARRLSGQPSAGLVPIH 106

Db 116 VAKAN--SLEPEWFFKNLSKDAEROLLAPGNTHGSLFIRESESTAGSFLSVRDFDQN 173

QY 107 VAKASPETLSQDPWFSGVSRTOAQQLLSPPNPGAFILRPSSSLGGYSLSVRA--- 162

Db 174 QGEVVKHYKIRNLNDGGFYI-SPRIITPGLHELVRHTNASDGLCTRLSRPCQTKPKP 232

QY 163 QAKVC-HYRVSMADGSLYQKRL-FPGLEELTYKANKWKLQNLPQPCMPKRA--P 218

Db 233 WDEWEVPRETLKLVRLGAGGFGVWGMGYNGHTKVAVKSLQGSMSDPAFLAEANLM 292

QY 219 R-QDWERPHSEFALGRKLGEGYFGEWGLWGLSLPVAIKYSANKKLTDLAKEIOTL 277

Db 293 KQIQHORLVLYAVTQ-EPIYITEYMENGLVDFLKTSGIKLTINKLMDMAAIAEG 351

QY 278 KGLRHERLIRLHVCSGGEPVIVTLMRKGNLQAFGLTPEGRALRPLPLLGACQVAEG 337

Db 352 MAFIERNYIHRDLRAANILVSDTLCKADIADFGARLIEDNEYTARGAKFIKWTAPEA 411

QY 338 MSYLEQVRVHRDLAARNVLVDGLACKVADFGARLLKDDIYSPSSSSKIPVKTAPEA 397

Db 412 INGTGTTIKSDVSGFILLTEIVTHGRIPYPGMTNPEVIONLERYRMVRDNCPEELYQ 471

QY 398 ANRVFSQKSDVSGFVLLHEVTYGCPEYEGMTNHTLQOIMRGYLRPAACPAEVIV 457

Db 472 LMRCKWERPEDPTFDYLRSLVLEDF 497

QY 458 LMLECRSSPEERPSFATLRKLHAI 483

RESULT 14

ENTRY B34104 #type complete

TITLE protein-tyrosine kinase (EC 2.7.1.112) src 2 - African clawed frog

ALTERNATE\_NAMES kinase-related transforming protein (src); kinase-related transforming protein (src) 2

ORGANISM #formal\_name Xenopus laevis #common\_name African clawed frog

DATE 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 20-Mar-1998

ACCESSIONS B34104; I51563

REFERENCE A34104

#authors Steele, R.E.; Unger, T.F.; Mardis, M.J.; Fero, J.B.

#journal J. Biol. Chem. (1989) 264:10649-10653

#title The two Xenopus laevis SRC genes are co-expressed and each produces functional pp(60src).

#cross-references MUID:89278134

#accession B34104

#status preliminary; not compared with conceptual translation

#molecule\_type mRNA

#residues 1-532 #label STE

#cross-references GB:M23422; GB:J04822; NID:g214796; PID:g214797

REFERENCE I51563

#authors Steele, R.E.

#journal J. Biol. Chem. (1989) 264:10649-10653

#title Nucleic Acids Res. (1985) 13:1747-1761

#title Two divergent cellular src genes are expressed in Xenopus laevis

#cross-references MUID:85215578

#accession I51563

#status preliminary; translated from GB/EMBL/DBJ

#molecule\_type DNA

#residues 439-492 #label ST2

#cross-references GB:M30858; NID:g214799; PID:g555569

GENETICS

#gene src

#introns 464/1

CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH3 homology

KEYWORDS ATP; autophosphorylation; phosphoprotein; phosphotransferase; tyrosine-specific protein kinase

FEATURE

87-136 #domain SH3 homology #label SH3\

147-244 #domain SH2 homology #label SH2\

264-522 #domain protein kinase homology #label KIN\

272-280 #region protein kinase ATP-binding motif

SUMMARY #length 532 #molecular-weight 59736 #checksum 7595

Query Match 36.3%; Score 1333; DB 2; Length 532;

Best Local Similarity 45.18; Pred. No. 3,62e-248;

Matches 195; Conservative 90; Mismatches 133; Indels 14; Gaps 10;

Db 85 FVALYDYESRTETDLSPKRGRLQIVNTEGDMWLARSLSGQT-GYIPSNVA-PS-DS 141

QY 56 FLALYDFTATRCGGELSVRRGRLCALEBEG-GGYIFARRLSGQPSAGLVPIHVAKASPT 114

Db 142 IQAEWILGKTRREARELLLSLENPRCTFLVRESEETKGAICLSVDYDASRGLNVKH 201

QY 115 LSDQFWFSGVSRTOAQQLLSPPNPGAFILRPSSSLGGYSLSVRA-QA-K--VCHY 169

Db 202 KIRLDSGGFYITRTQPSLSLQQLVAVYSKHADGICHLRTAVCPCTAKPQTQGLSKDAWEI 261

QY 170 RVSAADGSLYQKRLFPGLLEELTYKANKWKLQNLPQPC---MPQKAP-RODVWER 225

Db 262 PRDLRLLEKLGQCGFGVWGMGTNGTTRVAIKTLKPTGMSPEAFLOEAQVAKKLRLHEKL 321

QY 226 PHSEFALGRKLGEGYFGEWGLWGLSLPVAIKYSANKKLTDLAKEIOTLKLGRHERL 285

Db 322 VOLVAVSE-EPYIVITEYMENGLVDFLKGEMGRYLRDLPOLVDMAAQIASGMAYVERNN 380

QY 286 IRLHAVCSGGEPVIVTLMRKGNLQAFGLTPEGRALRPLPLLGACQVAEGMSYLEEQR 345

Db 381 YVHRDLRAANILVGENLVCKVADFGARLIEDNEYTARGAKFIKWTAPEAALYGRFTI 440

QY 346 VVHRDLAARNVLVDGLACKVADFGARLLKDDIYSPSSSSKIPVKTAPEAANRVFSQ 405

Db 441 KSDVMSFGILLTEITTKGRVPYPGMVNREVLDOVERGYRMPCCPPCESLHDMFOCNRK 500

QY 406 KSDVMSFGVLLHEVTYGCPEYEGMTNHTLQOIMRGYLRPAACPAEVIVLMCECWR 465

Db 501 DPEERPTFEYLQ 512

QY 466 SPEERPSFATLR 477

RESULT 15

ENTRY A34104 #type complete

TITLE protein-tyrosine kinase (EC 2.7.1.112) src 1 - African clawed frog

ALTERNATE\_NAMES kinase-related transforming protein (src); kinase-related transforming protein (src) 1

ORGANISM #formal\_name Xenopus laevis #common\_name African clawed frog

DATE 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 20-Mar-1998

ACCESSIONS A34104; I51564

REFERENCE A34104

#authors Steele, R.E.; Unger, T.F.; Mardis, M.J.; Fero, J.B.

#journal J. Biol. Chem. (1989) 264:10649-10653

Search completed: Thu May 20 12:21:21 1999  
Job time : 194 secs.

\*\*\*\*\*

WARRHEAD (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:21:39 1999; MasPar time 21.11 seconds  
Tabular output not generated. 653.617 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (1-488) from US09099053.pep (1 of 6)  
Perfect Score: 3671  
Sequence: 1 MEFFLRRRLAFLSFFWDKIW.....ERPSFATLREKLHAIHRCHP 488

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 50.552; Variance 83.305; scale 0.607

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2932	79.9	496	1	SRM_MOUSE TYROSINE-PROTEIN KINAS	0.00e+00
2	1442	39.3	506	1	SRK4_SPOLA TYROSINE-PROTEIN KINAS	0.00e+00
3	1391	37.9	505	1	SRK1_SPOLA TYROSINE-PROTEIN KINAS	0.00e+00
4	1388	37.8	507	1	LCK_CHICK PROTO-ONCOGENE TYROSIN	0.00e+00
5	1339	36.5	508	1	LCK_MOUSE PROTO-ONCOGENE TYROSIN	2.70e-294
6	1341	36.5	526	1	HCK_HUMAN TYROSINE-PROTEIN KINAS	8.39e-295
7	1336	36.4	529	1	FRK_HUMAN PROTO-ONCOGENE TYROSIN	1.56e-293
8	1332	36.3	505	1	FRK_HUMAN TYROSINE-PROTEIN KINAS	1.62e-292
9	1333	36.3	508	1	LCK_HUMAN PROTO-ONCOGENE TYROSIN	9.02e-293
10	1333	36.3	531	1	SRK2_XENLA TYROSINE-PROTEIN KINAS	9.02e-293
11	1329	36.2	531	1	SRK1_XENLA TYROSINE-PROTEIN KINAS	9.35e-292
12	1324	36.1	509	1	STK_HYDAT TYROSINE-PROTEIN KINAS	1.74e-290
13	1326	36.1	532	1	SRK_CHICK PROTO-ONCOGENE TYROSIN	5.40e-291
14	1326	36.1	568	1	SRK_AVISS TYROSINE-PROTEIN KINAS	5.40e-291
15	1321	36.0	503	1	HCK_RAT TYROSINE-PROTEIN KINAS	1.00e-289
16	1321	36.0	557	1	SRK_AVISS TYROSINE-PROTEIN KINAS	1.00e-289
17	1321	36.0	587	1	SRK_AVISS TYROSINE-PROTEIN KINAS	1.00e-289
18	1317	35.9	524	1	HCK_MOUSE TYROSINE-PROTEIN KINAS	1.04e-288
19	1315	35.8	536	1	SRK_AVISS TYROSINE-PROTEIN KINAS	3.34e-288
20	1315	35.8	535	1	BLK_HUMAN PROTO-ONCOGENE TYROSIN	3.34e-288
21	1303	35.5	504	1	BLK_HUMAN TYROSINE-PROTEIN KINAS	3.70e-285
22	1302	35.5	536	1	FYN_XIPHE PROTO-ONCOGENE TYROSIN	6.64e-285
23	1305	35.5	541	1	YES_CHICK PROTO-ONCOGENE TYROSIN	1.15e-285

24	1300	35.4	526	1	SRC_RSVH1 TYROSINE-PROTEIN KINAS	2.14e-284
25	1301	35.4	533	1	FYN_MOUSE PROTO-ONCOGENE TYROSIN	1.19e-284
26	1299	35.4	544	1	YES_XIPHE PROTO-ONCOGENE TYROSIN	3.83e-284
27	1296	35.3	526	1	SRC_RSVR TYROSINE-PROTEIN KINAS	2.21e-283
28	1293	35.2	526	1	SRC_RSVR TYROSINE-PROTEIN KINAS	1.27e-281
29	1287	35.1	537	1	YES_XENLA PROTO-ONCOGENE TYROSIN	4.23e-281
30	1284	35.0	543	1	YES_HUMAN PROTO-ONCOGENE TYROSIN	2.44e-280
31	1282	34.9	536	1	FYN_XENLA PROTO-ONCOGENE TYROSIN	7.83e-280
32	1281	34.9	536	1	FYN_HUMAN PROTO-ONCOGENE TYROSIN	1.40e-279
33	1282	34.9	541	1	YES_MOUSE PROTO-ONCOGENE TYROSIN	7.83e-280
34	1275	34.7	528	1	YES_AVISS TYROSINE-PROTEIN KINAS	4.66e-276
35	1266	34.5	511	1	LYN_RAT TYROSINE-PROTEIN KINAS	8.89e-276
36	1266	34.5	540	1	SCRN_MOUSE NEURONAL PROTO-ONCOGEN	8.89e-276
37	1262	34.4	498	1	BLK_MOUSE TYROSINE-PROTEIN KINAS	9.17e-275
38	1263	34.4	511	1	LYN_MOUSE TYROSINE-PROTEIN KINAS	5.12e-275
39	1263	34.4	533	1	FYN_CHICK PROTO-ONCOGENE TYROSIN	5.12e-275
40	1254	34.2	523	1	SRC_RSVA TYROSINE-PROTEIN KINAS	9.75e-273
41	1255	34.2	539	1	YES_CANFA PROTO-ONCOGENE TYROSIN	5.44e-273
42	1252	34.1	517	1	FGR_MOUSE PROTO-ONCOGENE TYROSIN	3.13e-272
43	1247	34.0	535	1	YRK_CHICK PROTO-ONCOGENE TYROSIN	5.78e-271
44	1237	33.7	511	1	LYN_HUMAN TYROSINE-PROTEIN KINAS	1.97e-268
45	1219	33.2	362	1	SRK2_SPOLA TYROSINE-PROTEIN KINAS	7.06e-264

ALIGNMENTS

RESULT 1  
ID SRM\_MOUSE STANDARD; PRT; 496 AA.  
AC Q62270; Q62360;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE SRM (EC 2.7.1.112) (PTK70).  
GN SRMS OR SRM.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN=C57BL/6; TISSUE=THYMUS;  
RX MEDLINE; 97369678.  
RA KAWACHI Y., NAKAUCHI H., OTSUKA F.;  
RT "Isolation of a cDNA encoding a tyrosine kinase expressed in murine skin.";  
RL EXP. DERMATOL. 21:533-538(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RX MEDLINE; 95021220.  
RA KOHMURA N., YAGI T., TOMOOKA Y., OYANAGI M., KOMINAMI R., TAKEDA N.,  
RT CHIBA J., IKAWA Y., AIZAWA S.;  
RA "A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption.";  
RL MOL. CELL. BIOL. 14:6915-6925(1994).  
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; D49427; G684972; -  
CC EMBL; D26186; G529073; -  
CC MGD; MGI:101865; SRMS.  
CC -----  
CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

```
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; PKINASE; 1.
DR HSPSP; P11362; IFGI.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 55 116 SH3.
FT DOMAIN 124 216 SH2.
FT NP_BIND 234 495 PROTEIN KINASE.
FT BINDING 262 262 ATP (BY SIMILARITY).
FT ACT_SITE 354 354 BY SIMILARITY.
FT MOD_RES 384 384 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 78 78 R -> G (IN REF. 2).
FT CONFLICT 236 238 LRK -> FGR (IN REF. 2).
FT CONFLICT 278 278 N -> I (IN REF. 2).
SQ SEQUENCE 496 AA; 55731 MW; FD4DEF6 CRC32;

Query Match 79.9%; Score 2932; DB 1; Length 496;
Best Local Similarity 78.2%; Pred. No. 0.00e+00;
Matches 383; Conservative 54; Mismatches 47; Indels 6; Gaps 4;

Db 1 MEPLRKLTLFSLFWDKIWPAD-ESEEDIPRIQGHNDNPVPEQAAVPEPCSFPPAPRL 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1 MEPLRRLRLAFLSFWDKIWPAGGPDHGTGOSLPDNDPVTIFA--EPCS-PFQ--L 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 60 FRALYDFTARCAEELUSVGRDRLYALKEGDYIFIAQRLSGPPSTGLVPVYLAATPEPP 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 FLALYDFTARGGELSVRGRDRLCALEGGGYIFARRLSGQSPAGLVPITHVAKASPEPL 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 120 SDQWYFSGISRAQAQQLLLSPANAPGAFILRPSESSIGGYSLVRAQAKVCHYRICAP 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 116 SDQWYFSGVSRTOAQQLLLSPNPPGAFILRPSESSIGGYSLVRAQAKVCHYRVSMA 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 180 SGSLYQGLQGLFSSDALLAYYKTNKLIQNPLLOPCIPQIPLVODEWERPSEVLRKK 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 176 DGSLYLQGRFLPGLLELLTYKANWKLIQNPLLOPCMPQKAPRODWERPSEFALGRK 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 240 LGEFFGVGEWGLWGLSIPVAVKVIKSDMKLADLTKEALKSLRHERLIRLHAICSIG 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 236 LGEYFGEWGEGLWGLSLPVAIKVKSANMKLTDLAKEIQTGLRHERLIRLHAVCSG 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 300 EPVIVITELMKGKGNLOVILGSSEGRKALSPLHLLGFACQVAGMSYLERRVVRDLAARN 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 296 EPVIVITELMRKGNLQAFGLTPEGRALPLPLGACQVAGMSYLEEQRVVVRDLAARN 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 360 VLVGDDLCKVADFGALRLKDDVYSPSSGSKI PVKWTAPAPEANRYVTSOKSDVWSFGIL 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 356 VLVDDGLACKVADFGALRLKDDIIVSPSSSKI PVKWTAPAPEANRYVTSOKSDVWSFGVL 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 420 LYEVFTYQCQPEYGMTNHETLQIISRGYRLPRPVCYAEVYVIMVCEWKGSPSEPTFAI 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 416 LHEVFTYQCQPEYGMTNHETLQIIMRGYRLPRPACPAEVYVIMLCEWRSSEPERSPAT 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 480 LREKLNAINR 489
|||||:|||||:
QY 476 LREKLHAIHR 485
|||||:|||||:

RESULT 2
ID SRK4_SPOLA STANDARD; PRT; 506 AA.
AC P42690;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRK4 (EC 2.7.1.112).
GN SRK4.
OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).
OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;
```

















```

RX MEDLINE; 85215578.
RA STEELE R.E.;
RT "Two divergent cellular src genes are expressed in Xenopus laevis.";
RL NUCLEIC ACIDS RES. 13:1747-1761(1985).
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M23422; G214797; -.
DR EMBL; M30858; G555569; -.
DR EMBL; M30857; G555569; JOINED.
DR PIR; B34104; B34104.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P00523; 2PTK.
DR TRANSFERASE; ATP-BINDING; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN
KW MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN
FT INIT_MET 0 BY SIMILARITY.
FT LIPID 1 MYRISTATE (BY SIMILARITY).
FT DOMAIN 79 140 SH2.
FT DOMAIN 146 243 SH3.
FT DOMAIN 265 518 PROTEIN KINASE.
FT NP_BIND 271 279 ATP (BY SIMILARITY).
FT BINDING 293 293 ATP (BY SIMILARITY).
FT ACT_SITE 384 384 ATP (BY SIMILARITY).
FT MOD_RES 414 414 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 531 AA; 59605 MW; C6ABE595 CRC32;

Query Match 36.3%; Score 1333; DB 1; Length 531;
Best Local Similarity 45.1%; Pred. No. 9.02e-293;
Matches 195; Conservative 90; Mismatches 133; Indels 14; Gaps 10;

Db 84 FVALYDYESRTEITDLSFRKGERLQIVNTEGDMWLARSLSGGT-GYIPSNVVA-PS-DS 140
Qy 56 FLALYDFTARCGGELSVERGRDLCALEG-GGYIFARLSGQPSAGLVPIITHVAKASPET 114

Db 141 IQABEWLKGITRREARLLSLENPRGTFVLVRESEITKGAYCLSVSDYASRGLNVKHY 200
Qy 115 LSDQPWYFSGVSRQAQQLLSPPNPGCAFLIRPSESSLGVSLSVA-QA-K---VCHY 169

Db 201 KIRKLDGGFITRTQPSFSSQQLVAYYSKADGLCHRLTAVCPATAQTOGLSKDAWEI 260
Qy 170 RVSMADGSLYLQGRLPFGLEELTYKANKWKLQNPLOPC---MPQKAP-RDYVWER 225

Db 261 PRDSIRLELKGQCGFGVWMTGNTTRVAIKTLKPCMTSPFAFLQEAQVMKMLRHEKL 320
Qy 226 PHSFALGRKLGEGYFGEVWGLWGLSLPVAIKVTKSANMKLTDIAKETQTLGRURHERL 285

Db 321 VOLXAVSE-BPIYIVTMYMSKGLDLDKGMGRYLRPLQIVDMAQAISGMAYVERMN 379
Qy 286 IRLHAVCSGGEPIVITELMRKGNLQALFIPTEGRALRPLPLGLFACQVAGMSYLEQR 345

Db 380 YVHRDLRAANILVGENLVCKVADFGCLARLIEDNEVTAQGAQKFKPKWTAPAEALYGRRTI 439
Qy 346 VVHRDLAARNVLVDGLACKVADFGCLARLIEDNEVTAQGAQKFKPKWTAPAEALYGRRTI 405
```

```

Db 440 KSDVWSFGILLTTLTKGRVPYPGMVNREVLDQVGRYMRPCPPDCPSLHDLMFQCWRK 499
Qy 406 KSDVWSFGVLLHVFYTCQCPYEGMTHETLQIMRGYLRPRAPCAEAVYVLMLECWRS 465

Db 500 DPEERTPEYLQ 511
Qy 466 SPEERPSFATLR 477

RESULT 11
ID SRC1_XENLA STANDARD; PRT; 531 AA.
AC P13115;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRC-1 (EC 2.7.1.112) (P60-SRC-1).
GN SRC-1.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP MEDLINE; 89278134.
RX STEELE R.E.; UNGER T.F.; MARDIS M.J.; FERRO J.B.;
RT "The two xenopus laevis SRC genes are co-expressed and each produces
RT functional pp60src."
RL J. BIOL. CHEM. 264:10649-10653(1989).
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M24704; G214805; -.
DR PIR; A34104; A34104.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P00523; 2PTK.
DR TRANSFERASE; ATP-BINDING; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN;
KW MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN
FT INIT_MET 0 BY SIMILARITY.
FT LIPID 1 MYRISTATE (BY SIMILARITY).
FT DOMAIN 79 140 SH2.
FT DOMAIN 146 243 SH3.
FT DOMAIN 265 518 PROTEIN KINASE.
FT NP_BIND 271 279 ATP (BY SIMILARITY).
FT BINDING 293 293 ATP (BY SIMILARITY).
FT ACT_SITE 384 384 BY SIMILARITY.
FT MOD_RES 414 414 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 531 AA; 59725 MW; AB0705D0 CRC32;

Query Match 36.2%; Score 1329; DB 1; Length 531;
Best Local Similarity 44.9%; Pred. No. 9.35e-292;
Matches 194; Conservative 92; Mismatches 132; Indels 14; Gaps 10;

Db 84 FVALYDYESRTEITDLSFRKGERLQIVNTEGDMWLARSLSGGT-GYIPSNVVA-PS-DS 140
Qy 56 FLALYDFTARCGGELSVERGRDLCALEG-GGYIFARLSGQPSAGLVPIITHVAKASPET 114
```

Db 141 IOAEWYLGKTRREARLLLSLENPRCTELVRESEETTKGAYCLSVSDYDANRGLNVKHY 200  
QY 115 LSDQFWIFSGVSRQAQQLLSPPNEPAGFLIRPSESSLGYSLSVRA-QA-K---VCHY 169  
Db 201 KIRLDSGGFYITSTQFSSLIQQLVAYYSKHADGSLCHRLTAVCPAKQTQGLSRDAWEI 260  
QY 170 RVMAADGSLYLQGRLPFGLELLTYKANKWLLQNPLQPC---MPQAP-RODVNER 225  
Db 261 PRDRLRLKLGQCGFGEVWNGTNGTTRVAIKTLKPGTMSPEAFLOQAQVMKKLRHEKL 320  
QY 226 PHSEFALGRKLGEYGEVWEGWGLGSLPVAIKVKSANMKLTDLAKEIQTLLGLRHERL 285  
Db 321 VOLYAVYSE-EPIYIVTYEYMSKGSLLDLFKGEMGRYLRLPOLVDMAAQIASGMAYVERMN 379  
QY 286 IRLHAVCGSGEPVIVITELMRKGNLQAFLTGPEGRALRLPPLLGFAQVAGMSYLEQR 345  
Db 380 YVHRDLRAANTLVGENLVCKVADFLGRLIEDNEYTAROGAKFPIKWTAPEAALYGRTI 439  
QY 346 VVHRDLARNVLVDGLACKVADFLGRLAKLDDIYSPSSSKIPVKTWAPEAANRVFSQ 405  
Db 440 KSDVMSFGILLTELTITKGRVPYPMGVNREVLDOVERGYRMPDPCPSLHDLMPQCWRK 499  
QY 406 KSDVMSFGVLLHEVETYGQCPYEGMTNHTLQIMRGYRLPRPACPAEYVYVLMLECHRS 465  
Db 500 DPERPTFEYIQ 511  
QY 466 SPEERPSFATLR 477

RESULT 12 STANDARD: PRT: 509 AA.  
ID STK\_HYDAT  
AC P17713;  
DT 01-AUG-1990 (REL. 15, CREATED)  
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE STK (EC 2.7.1.112) (P57-STK).  
GN STK.  
OS HYDRA ATTENUATA (HYDRA) (HYDRA VULGARIS).  
OC EUKARYOTA; METAZOA; CNIDARIA; HYDROZOA; HYDROIDA; ANTHOMEDUSAE;  
OC HYDRIDAE; HYDRA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90066418.  
RA BOSCH T.C.G., UNGER T.F., FISHER D.A., STEELE R.E.;  
RT "Structure and expression of STK, a src-related gene in the simple  
metazoan Hydra attenuata."  
RL MOL. CELL. BIOL. 9:4141-4151(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
PROTEIN TYROSINE PHOSPHATE.  
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC ENBL; M25245; G159274; -  
DR PIR; A34094; TVHAST.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50001; SH2; 1.  
DR PROSITE; PS50002; SH3; 1.  
DR PFAM; PF00017; SH2; 1.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00069; pkinase; 1.

HSP: P00523; 1PRL.  
KW TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; PHOSPHORYLATION;  
KW TRANSFERASE; ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN.  
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).  
FT DOMAIN 59 120 SH3.  
FT DOMAIN 126 218 SH2.  
FT DOMAIN 240 495 PROTEIN KINASE.  
FT BIND 246 254 ATP (BY SIMILARITY).  
FT BINDING 268 268 ATP (BY SIMILARITY).  
FT ACT\_SITE 360 360 BY SIMILARITY.  
FT MOD\_RES 390 390 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
SQ SEQUENCE 509 AA; 56885 MW; 2B724CE9 CRC32;

Query Match 36.1%; Score 1324; DB 1; Length 509;  
Best Local Similarity 43.5%; Pred. No. 1.74e-290;  
Matches 188; Conservative 90; Mismatches 144; Indels 10; Gaps 8;

Db 63 IFVALDYEARISEDLSPKKGRLQIINTADGDWYVARSLLTN-SEGYPSTIVVAPEKSY 121  
QY 55 LFALYDFTARGGELSVMRGDRLCAL-EGGGYIFARLSSGQPSAGLVPITHVAKASPE 113  
Db 122 E-AEE-WFGDVKRAEAEKRLMVRGLPSGTFILIRKAETAVGNFSLVRDGDVSKHYVRVK 179  
QY 114 TLSDQFWIFSGVSRQAQQLLSPPNEPAGFLIRPSESSLGYSLSVRAQAKVCHIRVSM 173  
Db 180 LDTGCFITTRAPFNSLXELVQHYTKDADGLV-CALTLPCPKDKPVTGGIAKDAWEIPRE 238  
QY 174 AADGSLYLQGRLPFGLELLTY-KANKWLLQNPLQPC---MPQAP-RODVNERPHS 228  
Db 239 SURLNRKLACQGFGEVWAGVWNTTQVAVKTLKPGTMSPASFLDEAGVMKKLRHKLVLQ 298  
QY 229 EFALGRKLGEYGEVWEGWGLGSLPVAIKVKSANMKLTDLAKEIQTLLGLRHERLRL 288  
Db 299 YAICSDREPIYIVTYEYMSKGSLLDLVLSRGEVNLQPLTIDMAAOVASCMAFLAQAQYIH 358  
QY 289 HAVCSGEPVIVITELMRKGNLQAFLTGPEGRALRLPPLLGFAQVAGMSYLEQRVVH 348  
Db 359 RDLAARNILVGENYICKVADFLGRLIEDDEYTAHEGAKFPIKWTAPEAALYNRFTIKSD 418  
QY 349 RDLAARNVLVDGLACKVADFLGRLAKLDDIYSPSSSKIPVKTWAPEAANRVFSQKSD 408  
Db 419 VMSFGILMAEIVTKGRIPYPMGNNAQTAEVEKGYRMPIMPGCPPELYNIMLQTNKDEP 478  
QY 409 VMSFGVLLHEVETYGQCPYEGMTNHTLQIMRGYRLPRPACPAEYVYVLMLECHRSPE 468  
Db 479 NRPTFDYLGQVL 490  
QY 469 ERPSFATLRKL 480

RESULT 13 SRC\_CHICK STANDARD: PRT: 532 AA.  
ID SRC\_CHICK  
AC P00523; Q91345; Q92013;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE SRC (EC 2.7.1.112) (P60-SRC).  
GN SRC.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 83155664.  
RA TAKEYA T., HANAFUSA H.;  
RT "Structure and sequence of the cellular gene homologous to the RSV  
src gene and the mechanism for generating the transforming virus."  
RL CELL 32:881-890(1983).  
RN [2]  
RP REVISION TO 525.  
RA TAKEYA T., HANAFUSA H.;  
RL CELL 34:319-319(1983).  
RN [3]



QY 286 IRLHAVSCGGEPIYIVTELNRKGNLQAFGLTPEGRALRLPPLILGFACQVAEGMSYLEEQR 345  
Db 381 YVHRDLRAANILVGENLVCKVADFGARLIEDNEYTAROGAKAPIKWTAPAAALYGRFTI 440  
QY 346 YVHRDLAARNVLVDGLACKVADFGARLLKDDIYSPSSSKIPVKTWTAPEAANYRVFSQ 405  
Db 441 KSDVMSFGILLTELTTKGRVPYPMVNRVLDQVERGYRMPCEPESLHDLMCOWRR 500  
QY 406 KSDVMSFVLLHEVFTYGCPEYEGMTNHTLQIMRGYRLPRAACPAEYVYVLMLECWRS 465  
Db 501 DPEERTFEYLQ 512  
QY 466 SPEERPSFATLR 477  
RESULT 14  
ID SRC\_AVISS STANDARD; PRT; 568 AA.  
AC P14084;  
DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE TRANSFORMING PROTEIN SRC (EC 2.7.1.112) (P60-  
SRC).  
GN V-SRC.  
OS AVIAN SARCOMA VIRUS (STRAIN S1).  
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87064539.  
RA IKAWA S., HAGINO-YAMAGISHI K., KAWA S., YAMAMOTO T., TOYOSHIMA K.;  
RT "Activation of the cellular src gene by transducing retrovirus.";  
RL MOL. CELL. BIOL. 6:2420-2428(1986).  
CC -!- FUNCTION: THIS PHOSPHOPROTEIN, REQUIRED FOR BOTH THE INITIATION  
CC AND THE MAINTENANCE OF NEOPLASTIC TRANSFORMATION, IS A PROTEIN  
CC KINASE THAT CATALYZES THE PHOSPHORYLATION OF TYROSINE RESIDUES  
CC IN VITRO.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
DR PIR: A25375; TVEVSL  
DR PROSITE: P500107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: P500109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: P500111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: P500001; SH2; 1.  
DR PROSITE: P500002; SH3; 1.  
DR PFAM: PF00017; SH2; 1.  
DR PFAM: PF00018; SH3; 1.  
DR PFAM: PF00069; pkinase; 1.  
DR HSSP: P00523; 2PTK.  
KW TYROSINE-PROTEIN KINASE; TRANSFORMING PROTEIN; ONCOGENE;  
KW TRANSFERASE; PHOSPHORYLATION; ATP-BINDING; MYRISTYLATION;  
KW SH3 DOMAIN; SH2 DOMAIN.  
FT LIPID 2 2 MYRISTATE.  
FT DOMAIN 81 142 SH3.  
FT DOMAIN 148 245 SH2.  
FT DOMAIN 267 520 PROTEIN KINASE.  
FT NP\_BIND 273 281 ATP (BY SIMILARITY).  
FT BINDING 295 295 ATP (BY SIMILARITY).  
FT ACT\_SITE 386 386 BY SIMILARITY.  
FT MOD\_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
SQ SEQUENCE 568 AA; 63632 MW; 7F080D52 CRC32;  
Query Match 36.1%; Score 1326; DB 1; Length 568;  
Best Local Similarity 44.4%; Pred. No. 5.40e-291;  
Matches 192; Conservative 92; Mismatches 134; Indels 14; Gaps 10;  
Db 86 FYALDYESTRTDLSFKKGERLQVNNTEGDNWLAHSLTTGOT-GYIPSNYYA-PS-DS 142  
QY 56 FLALDYFTARCGGELSVRRGDLRCALALEEG-GGYIFARRUSGQPSAGLVPITHAKASPET 114

Db 143 IQAEWEYFGKITRRESRLINPENPRGTFLVRESETHKAGYCLSVSDFDNARGLNVKHY 202  
QY 115 LSQDPWFYSGVSRTOAQOQLLLSPNPGAFILRSESLGGYSLSVRA-Q-AK---VCHY 169  
Db 203 KIRKLSGGYITRSTQSSLOQIVAYSKHADGLCHRLTNVCPTSKPTQOGLAKDAWEI 262  
QY 170 RVSMADSGSLYQKGRFLPGLEELLTYKANKWLIQNPLQPC---MPQKAP-RQDVWER 225  
Db 263 PRESRLRLEVLGQCGFVWMTGNGTTRVAIKTLKPGTMSPEAFLOEAQVYMKLRHEKL 322  
QY 226 PHSEFALGRKLGGYFGEVWEGWLGSLPVAIKVKSANKMLTDLAKEIQTLKGLRHERL 285  
Db 323 VRLYAVVSE-EPIYIVTEYMSKGLSLDLFLKMGCKYRLPOLVDMAAIAAGMAYVERMN 381  
QY 286 IRLHAVSCGGEPIYIVTELNRKGNLQAFGLTPEGRALRLPPLILGFACQVAEGMSYLEEQR 345  
Db 382 YVHRDLRAANILVGENLVCKVADFGARLIEDNEYTAROGAKAPIKWTAPAAALYGRFTI 441  
QY 346 YVHRDLAARNVLVDGLACKVADFGARLLKDDIYSPSSSKIPVKTWTAPEAANYRVFSQ 405  
Db 442 KSDVMSFGILLTELTTKGRVPYPMVNRVLDQVERGYRMPCEPESLHDLMCOWRR 501  
QY 406 KSDVMSFVLLHEVFTYGCPEYEGMTNHTLQIMRGYRLPRAACPAEYVYVLMLECWRS 465  
Db 502 DPEERTFEYLQ 513  
QY 466 SPEERPSFATLR 477  
RESULT 15  
ID HCK\_RAT STANDARD; PRT; 503 AA.  
AC P50545; Q64647;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE HCK (EC 2.7.1.112) (P56-HCK) (HEMOPHOETIC CELL  
DE KINASE).  
GN HCK.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92109719.  
RA OKANO Y., SUGIMOTO Y., FUKUOKA M., MATSUI A., NAGATA K.I., NOZAWA Y.;  
RT "Identification of rat cDNA encoding hck tyrosine kinase from  
RT megakaryocytes.";  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 181:1137-1144(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WISTAR; TISSUE-SPLEEN;  
RA VIJAYA GOURI B.S., REMA V., KAMATKAR S., SWARUP G.;  
RT "Nucleotide sequence of a cDNA coding for rat hck tyrosine kinase and  
RT characterization of its gene product.";  
RL J. BIOSCI. 19:117-129(1994).  
CC -!- FUNCTION: MAY SERVE AS PART OF A SIGNALING PATHWAY COUPLING THE FC  
CC RECEPTOR TO THE ACTIVATION OF THE RESPIRATORY BURST. MAY ALSO  
CC CONTRIBUTE TO NEUTROPHIL MIGRATION AND MAY REGULATE THE  
CC DEGRANULATION PROCESS OF NEUTROPHILS.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH MEMBRANES.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; S74141; G241437; -
DR EMBL; M83666; G204576; -
DR EMBL; X62345; G57582; -
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00019; SH3; 1.
DR HSP; P08631; 2HCK.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; ATP-BINDING;
KW MYRISTYLATION; SH2 DOMAIN; SH3 DOMAIN; ALTERNATIVE INITIATION.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
FT DOMAIN 55 115 SH3.
FT DOMAIN 121 218 SH2.
FT DOMAIN 239 492 PROTEIN KINASE.
FT NP_BIND 245 253 ATP (BY SIMILARITY).
FT BINDING 267 267 ATP (BY SIMILARITY).
FT ACT_SITE 358 358 BY SIMILARITY.
FT MOD_RES 388 388 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 51 51 F -> V (IN REF. 2).
FT CONFLICT 205 205 K -> R (IN REF. 2).
FT CONFLICT 306 306 I -> T (IN REF. 2).
SQ SEQUENCE 503 AA; 57016 MW; D94DE09 CRC32;

Query Match 36.0%; Score 1321; DB 1; Length 503;
Best Local Similarity 41.2%; Pred. No. 1.00e-289;
Matches 193; Conservative 114; Mismatches 144; Indels 17; Gaps 12;

Db 31 PDPTSPKKLGPN--INSLPGFVGESEDTIWAALDYEAHREDLSFKGDOMVVLLES 88
QY 26 PDHGTGSLDNDTDPVTLPAEPSCFPFQ-LFALYDFTARCGGELSVRGDRICALLEG 84
Db 89 GEWKARSLATK-KEGYIPSNYVARVN--SLETEEMFFKGISRKDAERHLLAPGNMLGSF 145
QY 85 GGYIFARRLSGQPSAGLVPIITHVAKASPETLSQDPWFSGVSRTOAQQLLLSPNPGAF 144
Db 146 MIRSETTKGYSLSVRDFDQHGDTVKHYKIRTLDSGFIYSPRSTFSSLOELVHYKK 205
QY 145 LIRPSESLGGYSLSVR--AQ-A-KVCHYRVSMADGSLXKQGRLPFGLEELTYKA 199
Db 206 GKDGICQK-LSVPCVSPKQPKWEKDAWEIPRESIQMEKKLGAGQFGEVWATYNKHTKV 264
QY 200 NWK-LIQNPLQPCMPQKA--PRQ-DWVERPHSEFALGKLGEGYGEVWEGWGLGSLPV 255
Db 265 AVKTMKPGSMVSEAFLEAANLMKTLQHKVLKHAHVVSQ-EPFIVTFEMAKGSLDLFLK 323
QY 256 AIKVIKSANMKLTDLAKETIQLKGLRHERLRLHAVCSGGEPVIVTELMRKGNLQAFIG 315
Db 324 SEEGSQPLPKLIDFSAQISGMAFIEQRYIHRDLRAANILVSALVCKTADFGLARI 383
QY 316 TPEGRLALPPLGFAQVAGESMSYLEEQRYVVRDLAARNVLVDGGLACKVADFGLARLL 375
Db 384 EDNEYTAREGAKFPKKTAPAINFGSTIKSDWWSFGILLMEIVTYGRIPIYPMGMSNPEV 443
QY 376 KDDIYSPSSSSKIPVKWTAPAPAAVRYVFSQSDWWSFGVLLHEVITYGQCPYEGMTNHET 435
Db 444 ITRALEHYRMPRPDNCPEELYSIMIRCKNRPERPTFEYIQSVLDDF 491
QY 436 LQIMRGYRLPPAACPAEVVYVLMLECWRSPEERPSFATLREKLHAI 483
```

Search completed: Thu May 20 12:22:23 1999  
Job time : 44 secs.



```

MPsrch_pp      protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 20 12:22:43 1999; MasPar time 45:53 Seconds
584.987 Million cell updates/sec

```

ERPSEATLREKLHAIHRCHP 488

Statistics: Mean 48.824; Variance 91.236; scale 0.535

## SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description	Pred. No.
1	1437	39.1	517	5	Q94879	DSRC41.	2.78e-382
2	1366	37.2	451	11	Q64334	SRC-RELATED INTESTINAL TYROSINE KINASE.	2.47e-366
3	1355	36.9	451	13	Q13882	NON-RECEPTOR PROTEIN T	7.27e-364
4	1342	36.6	496	13	Q93411	SRC RELATED TYROSINE K	6.00e-361
5	1335	36.4	506	11	Q62562	SRC RELATED TYROSINE K	2.23e-359
6	1335	36.4	512	11	Q61364	B-CELL SRC-HOMOLOGY TY	2.23e-359
7	1334	36.3	512	11	Q61745	B-CELL SRC-HOMOLOGY TY	3.73e-359
8	1321	36.0	533	13	Q98915	GENE C-SRC PRODUCING P	3.07e-356
9	1321	36.0	533	13	Q90992	C-SRC.	3.07e-356
10	1315	35.8	587	14	Q64917	PROTEIN-TYROSINE KINAS	6.78e-355
11	1307	35.6	488	13	Q13064	LYN PROTEIN TYROSINE K	4.21e-353
12	1306	35.6	526	11	Q60867	H-19 PROVIRAL SEQUENCE	7.05e-353
13	1304	35.5	512	4	Q12850	LYMPHOCTE-SPECIFIC PR	1.98e-352
14	1300	35.4	505	4	Q16291	BLK=PROTEIN TYROSINE K	1.56e-351
15	1299	35.4	534	4	Q16248	P59FYN.	2.61e-351
16	1296	35.3	517	5	Q77050	SRC-TYPE PROTEIN TYROS	1.23e-350
17	1296	35.3	525	14	Q92806	P60 SRC.	1.23e-350
18	1291	35.2	526	14	Q07461	TYROSINE-PROTEIN KINAS	1.62e-349
19	1281	34.9	527	13	Q91952	C-SRC TYROSINE KINASE.	2.80e-347
20	1281	34.9	537	11	Q62344	PROTO-ONCOGENE FYN.	2.80e-347

```

Qy 212 CMPQK-APR-QDVWRPHSEFALGKLGEGYGEYEWGLWGLSLPVAIKVKSANMKLTD 269
Db 288 FLAEAIQMKKLRHTKLIQIYAVCTVEEPIIITELTMKHSLSLEYLOAIAGKRSKMQTL 347
Qy 270 LAKEIQTGLKGLRHLRLHRAVCSGEPYIIVTELMRKGNLQAF-LGTPE-GRALRLPL 327
Db 348 IDMAQIAAGMAYLSEYKVIHRDLAARNVLVGDGNIVKIADFGLARLIKEDYEYARVGAR 407
Qy 328 LGFACQAGMSYLBEEQVRVHDLAARNVLVDGGLACKVADFGLARLLKDDIYSPSSSK 387
Db 408 FPIKWTAPAAANSFSSIKSDWSFGILITELVTYGRIPYGMNTAEVLTOVQEHGYRMPQ 467
Qy 388 IPVKWTAPAAANYRVSQKSDWSFGVLLHEVFTYGCPEYEGTNNHETLQIMRGYRLPR 447
Db 468 PNPCEPRIYEIMLECHWKDPMRRTFTLQWKLEDF 503
Qy 448 PAACPAEVVLMLECWSSPERPSPFATLREKLHAI 483

RESULT 2
ID Q64434 PRELIMINARY; PRT; 451 AA.
AC Q64434;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SRC-RELATED INTESTINAL KINASE (EC 2.7.1.112)
DE (PROTEIN-TYROSINE KINASE) (TYROSYLPROTEIN KINASE) (PROTEIN KINASE
DE (TYROSINE)) (HYDROXYARYL-PROTEIN KINASE).
GN SIK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCUTOGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HSD: ICR AND BALB/C; TISSUE=INTESTINE;
RX MEDLINE; 95140424.
RA VASIOUKHIN V., SERFAS M.S., SIYANOVA E.Y., POLONSKAIA M.,
RA COSTIGAN V.J., LIU B., THOMASON A., TYNER A.L.;
RT "A novel intracellular epithelial cell tyrosine kinase is expressed
in the skin and gastrointestinal tract.;"
RL ONCOGENE 10:349-357(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HSD: ICR; TISSUE=INTESTINE;
RX MEDLINE; 94268846.
RA SIYANOVA E.Y., SERFAS M.S., MAZO I.A., TYNER A.L.;
RT "Tyrosine kinase gene expression in the mouse small intestine.;"
RL ONCOGENE 9:2053-2057(1994);
RN [3]
RP SEQUENCE OF 1-77 FROM N.A.
RC STRAIN=BALB/C;
RA SIYANOVA E.Y.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
CC TYROSINE PHOSPHATE.
DR EMBL; U16805; G847795; -.
DR EMBL; AF016545; G2738777; -.
DR MGD; MGI-99683; SIK.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
KW TRANSFERASE.
SQ SEQUENCE 451 AA; 51972 MW; 5A749D95 CRC32;

Query Match 37.2%; Score 1366; DB 11; Length 451;
Best Local Similarity 46.8%; Pred. No. 2.47e-266;
Matches 204; Conservative 74; Mismatches 150; Indels 8; Gaps 8;

Db 13 YVGLWDFKARTDEELSFQAGDLHLHTKKEELWWATLLDAEGKALAEGVVPHNYLAEKET 72
Qy 56 FLALYDFTRCGGELSVRGDRLEALEGGGYIFARRLSGQPSAGLVP-ITHVAKASPET 114

```

```

Db 73 VSEPFWFCICISRSAMHRLQEDNSKGAFLIRVSOKPGADYVLSYRDAQAVPHYRIWKN 132
Qy 115 LSDQWYFSGVSRTOAQQLLPPNEPAGFLIRPSESSLGYSLSYRAQAKVCHYRVSMA 174
Db 133 NGRUHLNEAVSFNSLSELDVYHKTO-SLSHGLQLSMPCKWKHTEPLPHWDDWDERPREF 191
Qy 175 ADGSYLQKGRFLPGLELLTYKANKLIQN-PLLOCPMPKA-P-RQ-DWVERPHSEF 230
Db 192 TLCKLKGAGYGEFEALWKQGVHVAVKVISRDNLHHTFOAEIOAMKLRHKHLSLY 251
Qy 231 ALGRKLGEYGEYEWGLWGLSLPVAIKVKSANM-KLTDLAKEIOTLGLRHLRLH 289
Db 252 AVATAGDPVIITELMPKGNLLOLLRDSDEKALPILELVDFASQVAGCMCYLESQYIHR 311
Qy 290 AVCSGEPYIIVTELMRKGNLQAF-LGTPEGRALRLPLPGLGFACQVAGMSYLEEQRVHR 349
Db 312 DLAAARNVLVTENNLCKVGDGFLARLVKEDIY-LSHEHNVPYKWTAPALSRGHYSKSDV 370
Qy 350 DLAAARNVLVDGGLACKVADFGLARLLKDDIYSPSSSKIPVKWTAPAAANYRVSQKSDV 409
Db 371 WSFGVLLHEIFSRGOMPYPGMSNHETFLVDAGYRMPCEPPIHKLMLSCWSRDPKQ 430
Qy 410 WSFGVLLHEVFTYGCPEYEGTNNHETLQIMRGYRLPRPAACPAEVVLMLECWSSPEE 469
Db 431 RCFKDKLCEKLGITR 446
Qy 470 RPSFATLREKLHAIHR 485

RESULT 3
ID Q13882 PRELIMINARY; PRT; 451 AA.
AC Q13882;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE TYROSINE KINASE.
GN BRK OR PTK6.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST CARCINOMA;
RX MEDLINE; 94309916.
RA MITCHELL P.J., BARKER K.T., MARTINDALE J.E.;
RT "Cloning and characterisation of cDNAs encoding a novel non-receptor
tyrosine kinase, brk, expressed in human breast tumours.;"
RL ONCOGENE 9:2383-2390(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97430836.
RA PARK S.H., LEE K.H., KIM H., LEE S.T.;
RT "Assignment of the human PTK6 gene encoding a non-receptor protein
tyrosine kinase to 20q13.3 by fluorescence in situ hybridization.;"
RL CYTOGENET. CELL GENET. 77:271-272(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98419955.
RA LEE H., KIM M., LEE K.-H., KANG K.-N., LEE S.-T.;
RT "Exon-intron structure of the human PTK6 gene demonstrates that PTK6
RT constitutes a distinct family of non-receptor tyrosine kinase.;"
RL MOL. CELLS 8:401-407(1998).
DR EMBL; X78549; G515026; -.
DR EMBL; U61412; G3551753; -.
DR EMBL; U61406; G3551753; JOINED.
DR EMBL; U61407; G3551753; JOINED.
DR EMBL; U61408; G3551753; JOINED.
DR EMBL; U61409; G3551753; JOINED.
DR EMBL; U61410; G3551753; JOINED.
DR EMBL; U61411; G3551753; JOINED.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

```



```
Db 157 VLDEGVVHYRIRRLDEGGFFLTKRKTFTSTINEFVNYYTTSDGLCVKLEKPKIQVPT 216
QY 160 VRAQAKVHYRVSMAADGSLYQKGRFLPGLLELLTYKANKW-L-I--QNP-L-LQPCM 213
Db 217 PFDLSKTVQDWEIDRNSIOLLKRLSGSGQFGEVWEGWLNNTTPVAVKTLKPGSDPNDEL 276
QY 214 PQKAP-RQ-DVWERPHSEFALGRKLGEGYFGEVWEGWLNNTTPVAVKTLKPGSDPNDEL 271
Db 277 REAQIMKSLRHPKLIQIYAVCTLEDPIYIITELMRHSGSQEYQLQNDGSGIRLITQQYDMA 336
QY 272 KEIOTLGLRHERLIRLHVCSSGEPYIVITELMRKGNLQAFGLTPEGRALRPLPLGFA 331
Db 337 AQVAGMAYLESQNYIHRDLAARNVLGHEHNIYKVADFGARVFKVDNEDIYKSKHKL 396
QY 332 CQVAGMSYLEEQRVVHRDLAARNVLVDGGLACKVADFGARLLK-D--DIYSPSSSKI 388
Db 397 PVKWTAPAEARTNKFSKSDVWSFGILLYEITYGKMPYSGMTGAQVIMLGQNYRLPOP 456
QY 389 PVKWTAPAEANRYRFSQKSDVWSFGVLLHEVFTYGQCPYEGMTNHETLQOIMRGYLRPR 448
Db 457 SNGPEQFYSIMMECNVPEKQRTFTFELHWKL 488
QY 449 AACPAEYVLMLECRSSPEERSFATLREKL 480

RESULT 6
ID Q61364 PRELIMINARY; PRT; ...512 AA...
AC Q61364;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE B-CELL SRC-HOMOLOGY TYROSINE KINASE (PROTEIN TYROSINE KINASE).
GN FRK OR BSK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95137395.
RA OBERG-WELSH C., WELSH M.;
RT "Cloning of BSK, a murine FRK homologue with a specific pattern of
   tissue distribution.";
RL GENE 152:239-242(1995).
DR EMBL; L36132; G77773; -.
DR MGD; MGI:103265; FRK.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PFAM; PF000017; SH2; 1.
DR PFAM; PF000018; SH3; 1.
DR PFAM; PF000069; pkinase; 1.
SQ SEQUENCE 512 AA; 5891 MW; DEC53C7 CRC32;

Query Match 36.48; Score 1335; DB 11; Length 512;
Best Local Similarity 43.28; Pred. No. 2.23e-259;
Matches 198; Conservative 102; Mismatches 140; Indels 18; Gaps 16;

Db 38 PEAPRSQEPERSHGQYFVALDFOARTAEPLSFRAQDKQLVLDTSHEGWLAHLE-KKG 96
QY 40 PVTPLPAEPCSPFPQLFLALYDFTARCGLSVRRGDRCLALEEG-GGYIFARLSGQPS 98

Db 97 TGLGQQLQGYIPSNYVAEDRSLOAEPWFFGAIKRADAEKQLLYSENQTGAFLRISESQK 156
QY 99 AGLVP-I-THV-AK-ASPE-TLSDQPYFSGVSRTOAQQLLSPNPEGAFLRIPSESL 153

Db 157 GDFSLSVLDGSGVVKHYRIRLDEGGFFLTRKRVSTNEFVNYYTTSDGLCVKLEKPC 216
QY 154 GGYSLVRAQAKVCHYRVSMAADGSLYQKGRFLPGLLELLTYKANKW-L-I--QNP-L 208

Db 217 TGLGQQLQGYIPSNYVAEDRSLOAEPWFFGAIKRADAEKQLLYSENQTGAFLRISESQK 156
QY 99 AGLVP-I-THV-AK-ASPE-TLSDQPYFSGVSRTOAQQLLSPNPEGAFLRIPSESL 153

Db 157 GDFSLSVLDGSGVVKHYRIRLDEGGFFLTRKRVSTNEFVNYYTTSDGLCVKLEKPC 216
QY 154 GGYSLVRAQAKVCHYRVSMAADGSLYQKGRFLPGLLELLTYKANKW-L-I--QNP-L 208

Db 217 KIQVPTPFDLSYKTADQWEIDRNSIQLLKRGLSGQFGEVWEGWLNNTTPVAVKTLKPGSM 276
QY 209 -LQPCMPQKAP-R-QDWERPHSEFALGRKLGEGYFGEVWEGWLNNTTPVAVKILKSANM 265

Db 277 DPNDFLBEAQIMKSLRHPKLIQIYAVCTLEDPIYIITELMRHSGSQEYQLQNDGSGKIHI 336
QY 266 KLTDLAKEIOTLGLRHERLIRLHVCSSGEPYIVITELMRKGNLQAFGLTPEGRALR 325

Db 337 QQVDMAAQVAGMAYLESQNYIHRDLAARNVLGHEHNIYKVADFGARVFKVDNEDIY 396
QY 326 PLGFAQVAGMAYLESQNYIHRDLAARNVLVDGGLACKVADFGARLLK-D--DIYSP 382

Db 397 KHEIKLPVKWTAPAEARTNKFSKSDVWSFGILLYEITYGKMPYSGMTGAQVIMLGQ 456
QY 383 SSSSKIPVKWTAPAEANRYRFSQKSDVWSFGVLLHEVFTYGQCPYEGMTNHETLQOIM 442

Db 457 YRLPOPENCOFQFYSIMMECNVPEKQRTFTFELHWKL 494
QY 443 YRLPRPAACPAEYVLMLECRSSPEERSFATLREKL 480
```

```
Db 277 DPNDFLBEAQIMKSLRHPKLIQIYAVCTLEDPIYIITELMRHSGSQEYQLQNDGSGKIHI 336
QY 266 KLTDLAKEIOTLGLRHERLIRLHVCSSGEPYIVITELMRKGNLQAFGLTPEGRALR 325
Db 337 QQVDMAAQVAGMAYLESQNYIHRDLAARNVLGHEHNIYKVADFGARVFKVDNEDIY 396
QY 326 PLGFAQVAGMAYLESQNYIHRDLAARNVLVDGGLACKVADFGARLLK-D--DIYSP 382
Db 397 KHEIKLPVKWTAPAEARTNKFSKSDVWSFGILLYEITYGKMPYSGMTGAQVIMLGQ 456
QY 383 SSSSKIPVKWTAPAEANRYRFSQKSDVWSFGVLLHEVFTYGQCPYEGMTNHETLQOIM 442
Db 457 YRLPOPENCOFQFYSIMMECNVPEKQRTFTFELHWKL 494
QY 443 YRLPRPAACPAEYVLMLECRSSPEERSFATLREKL 480

RESULT 7
ID Q61745 PRELIMINARY; PRT; 512 AA.
AC Q61745;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE B-CELL SRC-HOMOLOGY TYROSINE KINASE (INTESTINAL TYROSINE KINASE).
GN FRK OR IYK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE; 95251656.
RA THUVESON M., ALBRECHT D., ZUERCHER G., ANDRES A., ZIEMIECKI A.;
RT "Iyk, a novel intracellular protein tyrosine kinase differentially
   expressed in the mouse mammary gland and intestine.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 209:582-589(1995).
DR EMBL; Z48757; G736264; -.
DR MGD; MGI:103265; FRK.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PFAM; PF000017; SH2; 1.
DR PFAM; PF000018; SH3; 1.
DR PFAM; PF000069; pkinase; 1.
SQ SEQUENCE 512 AA; 58928 MW; 808D1612 CRC32;

Query Match 36.38; Score 1334; DB 11; Length 512;
Best Local Similarity 43.08; Pred. No. 3.73e-259;
Matches 197; Conservative 103; Mismatches 140; Indels 18; Gaps 16;

Db 38 PEAPRSQEPERSHGQYFVALDFOARTAEPLSFRAQDKQLVLDTSHEGWLAHLE-KKG 96
QY 40 PVTPLPAEPCSPFPQLFLALYDFTARCGLSVRRGDRCLALEEG-GGYIFARLSGQPS 98

Db 97 TGLGQQLQGYIPSNYVAEDRSLOAEPWFFGAIKRADAEKQLLYSENQTGAFLRISE 156
QY 99 AGLVP-I-THV-AK-ASPE-TLSDQPYFSGVSRTOAQQLLSPNPEGAFLRIPSESL 153

Db 157 GDFSLSVLDGSGVVKHYRIRLDEGGFFLTRKRVSTNEFVNYYTTSDGLCVKLEKPC 216
QY 154 GGYSLVRAQAKVCHYRVSMAADGSLYQKGRFLPGLLELLTYKANKW-L-I--QNP-L 208

Db 217 KIQVPTPFDLSYKTADQWEIDRNSIQLLKRGLSGQFGEVWEGWLNNTTPVAVKTLKPGSM 276
QY 209 -LQPCMPQKAP-R-QDWERPHSEFALGRKLGEGYFGEVWEGWLNNTTPVAVKILKSANM 265

Db 277 DPNDFLBEAQIMKSLRHPKLIQIYAVCTLEDPIYIITELMRHSGSQEYQLQNDGSGKIHI 336
QY 266 KLTDLAKEIOTLGLRHERLIRLHVCSSGEPYIVITELMRKGNLQAFGLTPEGRALR 325

Db 337 QQVDMAAQVAGMAYLESQNYIHRDLAARNVLGHEHNIYKVADFGARVFKVDNEDIY 396
QY 326 PLGFAQVAGMAYLESQNYIHRDLAARNVLVDGGLACKVADFGARLLK-D--DIYSP 382
```

```
Db 397 KHEIKLPVKWTAPEAIRTNKFSIKSDVMSFGILLYEITYYKMPYSQMTGAQVIMQLSON 456
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 383 SSSKIPVKWTAPEAIRNYRVSQSDVMSFGVLLHEVFTYGCPEYEGMTNHTLQIMRG 442
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 457 YRLPQSNCPQOQYSIMLECNWVPEKQPTFFELHWKL 494
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 443 YRLPRAACPAEYVYVLMLEWCRSPSPERPSFATLREKL 480
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||

RESULT 8
ID Q98015 PRELIMINARY; PRT; 533 AA.
AC Q98915; Q91343;
DT 01-FEB-1997 (TREMBREL. 02, CREATED)
DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE GENE C-SRC PRODUCING PROTEIN PF60-C-SRC.
DE THIS GENE IS HOMOLOGOUS TO THE ROUS SARCOMA VIRUS GENE V-SRC
DE (PF60C-SRC).
GN C-SRC OR C-SCR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83155664.
RA TAKEYA T., HANAFUSA H.;
RT "Structure and sequence of the cellular gene homologous to the RSV
RT src gene and the mechanism for generating the transforming virus.";
RL CELL 32:881-890(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97008971.
RA WEILAND A., NEUBAUER G., COURTREIDGE S.A., MANN M., WIENERGA R.;
RT "The purification and characterization of the catalytic domain of Src
RT expressed in Schizosaccharomyces pombe. Comparison of
RT unphosphorylated and tyrosine phosphorylated species.";
RL EUR. J. BIOCHEM. 240:756-764(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA WEILAND A.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 484-533 FROM N.A.
RX MEDLINE; 91304409.
RA DORAI T., LEVY J.B., KANG L., BRUGGE J.S., WANG L.H.;
RT "Analysis of cDNAs of the proto-oncogene c-src: heterogeneity in 5'
RT exons and possible mechanism for the genesis of the 3' end of
RT v-src.";
RL MOL. CELL. BIOL. 11:4165-4176(1991).
DR EMBL; V00402; E281134; -.
DR EMBL; V00402; E181088; -.
DR EMBL; S43579; E97011; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 533 AA; 60010 MW; 0D446FF3 CRC32;

Query Match 36.0%; Score 1321; DB 13; Length 533;
Best Local Similarity 44.4%; Pred. No. 3.07e-256;
Matches 192; Conservative 92; Mismatches 134; Indels 14; Gaps 10;

Db 86 FVALDYESTRTDLSFKKGERLQIVNNTGDMWLASHLTGTGT-GYIPSNVA-PS-DS 142
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 56 FLALYDFTARCGLSGELSVRRGDRLCALLEG-GGYIFARRLSQPSAGLVPIITHVAKASPET 114
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 143 IQAEWYFGKITRRESERLLNPNPRGTFLVRESEITTKGAYCLSVSDFDNAGLNKHY 202
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 115 LSDQPWVFSVSRQAQQLLSPNPPGAFILRPSESSLGGSLSVRA-Q-AK---VCHY 169
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 203 KIRKLDGGFYITSRTOFSSLOQLVAYYSKHADGLCHRLTNVCPSTKPTQGLAKDAWEI 262
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 170 RVSMADGSLYQKRGFLPGLLELITYYKANWKLQNPLQPC---MPQKAP-RODVWER 225
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 263 PRESRLRLEVLKGOCFCFGEVWMTNGTTRVAIKTLKPGTMSPEAFLOEAQVNMKLRHEKL 322
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 226 PHSEFALGRKLGEYFGVEWGLWGLSLPVAIKVKSANMKLTDLAKETIQLKGLURHERL 285
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
```

```
QY 170 RVSMADGSLYQKRGFLPGLLELITYYKANWKLQNPLQPC---MPQKAP-RODVWER 225
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 263 PRESRLRLEVLKGOCFCFGEVWMTNGTTRVAIKTLKPGTMSPEAFLOEAQVNMKLRHEKL 322
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 226 PHSEFALGRKLGEYFGVEWGLWGLSLPVAIKVKSANMKLTDLAKETIQLKGLURHERL 285
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 323 VOLYAVYSE-EPIYIVTEYMSKGLDPLKGMGKYLRLPOLVDNMAAQIASGMAYVERMN 381
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 286 IRLHAVCSGGEPVIVTELMRKGNLQAFLTGPEGRALRLPLLGFAQCQVAGMSYLEQR 345
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 382 YVHRDLRAANTLVGENLVCKVADFLGRLIEDNEYTARQGAFFIKKTAPEAALGRRTI 441
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 346 VVHRDLAARNVLDDGLACKVADEGLARLLDKDDIYSPSSSKIPYKWTAPAANAIVRFSQ 405
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 442 KSDVMSFGILLTTLTKGRVPPGMVNVREVLDOVERGYRMPCCPPCESLHDLMCQCRK 501
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 406 KSDVMSFGVLLHEVFTYGCPEYEGMTNHTLQIMRGYRLPRPAACPAEYVYVLMLEWCRS 465
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 502 DPEERPTFEYQLQ 513
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 466 SPEERPSFATLR 477
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||

RESULT 9
ID Q90992 PRELIMINARY; PRT; 533 AA.
AC Q90992;
DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE C-SRC.
GN SRC.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA HANAFUSA H.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83155664.
RA TAKEYA T., HANAFUSA H.;
RT "Structure and sequence of the cellular gene homologous to the RSV
RT src gene and the mechanism for generating the transforming virus.";
RL CELL 32:881-890(1983).
DR EMBL; J00844; G901820; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 533 AA; 60038 MW; 8B987D6B CRC32;

Query Match 36.0%; Score 1321; DB 13; Length 533;
Best Local Similarity 44.4%; Pred. No. 3.07e-256;
Matches 192; Conservative 92; Mismatches 134; Indels 14; Gaps 10;

Db 86 FVALDYESTRTDLSFKKGERLQIVNNTGDMWLASHLTGTGT-GYIPSNVA-PS-DS 142
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 56 FLALYDFTARCGLSGELSVRRGDRLCALLEG-GGYIFARRLSQPSAGLVPIITHVAKASPET 114
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 143 IQAEWYFGKITRRESERLLNPNPRGTFLVRESEITTKGAYCLSVSDFDNAGLNKHY 202
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 115 LSDQPWVFSVSRQAQQLLSPNPPGAFILRPSESSLGGSLSVRA-Q-AK---VCHY 169
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 203 KIRKLDGGFYITSRTOFSSLOQLVAYYSKHADGLCHRLTNVCPSTKPTQGLAKDAWEI 262
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 170 RVSMADGSLYQKRGFLPGLLELITYYKANWKLQNPLQPC---MPQKAP-RODVWER 225
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
Db 263 PRESRLRLEVLKGOCFCFGEVWMTNGTTRVAIKTLKPGTMSPEAFLOEAQVNMKLRHEKL 322
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
QY 226 PHSEFALGRKLGEYFGVEWGLWGLSLPVAIKVKSANMKLTDLAKETIQLKGLURHERL 285
      1:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:||||| 11:|||||
```





Query Match 35.4%; Score 1300; DB 4; Length 505;  
Best Local Similarity 44.3%; Pred. No. 1,566-251;  
Matches 195; Conservative 92; Mismatches 136; Indels 17; Gaps 14;

Db 61 HFVALYDYTAMNDRDLOMLKGEKLVQKGTGDMWLARSVLT-GREGYVPSNFVARV--E 117  
Qy 54 QFLALYDFTARCAGGELSYYRRDRDLCALCEGGYIFARRLSGOPSAGLVPIITHAKASPE 113  
Db 118 SLEMERWFRSOGKAEARQLLAPINKAGSFLIRSETNKGAFSLSVKDVTTQGELIKHY 177  
Qy 114 TLDSPQWFGSVRTOAQOALLSPNEPQAFILRSESSLGYSLSVR--AAQKVC-HY 169  
Db 178 KTRCIDEGYI-SPRTIFPSQALVQHYSKKGDLQCR-LTLPQVRPAPQNPWAQDEWE 235  
Qy 170 RVSMAADGSLYLQKGR-LFPGLEELTY-KANWKLIQNPLQPCN-P-QKAPR-QDVWE 224  
Db 236 IPQSLRLVRKLGSGQFGEVWNGYKNNMKVAIKTLKEGTMSPFAFLGEANYMKALQHER 295  
Qy 235 RPHEFALGRKLGEYFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIQTLKGLRHER 284  
Db 296 LVRLYAVVTK-EPIYIVTEYMAARGCLLDLKTDEGSRLSLPRLIDMSAOIAEGMAYIERM 354  
Qy 285 LIRLHAVCSGGEPIVITELMRKGNLQAFLTGPEGRALRPLLLGFACQVAGMSYLEEQ 344  
Db 355 NSIHRDLRAANILVSEALCKIADFLGARIIDSE-YTAQEGAKFPKWTAPAYHFGVET 413  
Qy 345 RVVHRDLAARNVLVDGLACKVADFLGRLKDDIYSPSSSKIPVKTWTAPEAANYRVFS 404  
Db 414 IKADVWSFGVLLMEVTVGRVPYPGNSNENLGRYMRPDTCPELYRGVTAECW 473  
Qy 405 QKSDVMSFGVLLHEVFTYGCQPYEGMTNHTLQQIMRGYRLPRPAACPAEVY-VLMLECW 463  
Db 474 RSRPEERTFEFLQSVLEDF 493  
Qy 464 RSSPEERPSFATLRKLHAI 483

RESULT 15  
ID Q16248 PRELIMINARY; PRT; 534 AA.  
AC Q16248;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE P59fyn.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95123064.  
RA RIGLEY K., SLOCOMBE P., PROUDFOOT K., WAHID S., MANDAIR K.,  
RA BEBBINGTON C.  
RT "Human p59fyn(T) regulates OKT3-induced calcium influx by a mechanism  
RT distinct from PIP2 hydrolysis in Jurkat T cells.";  
RL J. IMMUNOL. 154:1136-1145(1995).  
DR EMBL; S74774; G802051; -  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PFAM; PF00017; SH2; 1.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00069; pkinase; 1.  
SQ SEQUENCE 534 AA; 60226 MW; D3941652 CRC32;

Query Match 35.4%; Score 1299; DB 4; Length 534;  
Best Local Similarity 43.4%; Pred. No. 2,616-251;  
Matches 188; Conservative 94; Mismatches 137; Indels 14; Gaps 12;

Db 86 LFVALYDYDEARTDDLSFHKGEKQLNSGSDWNEARSLTT-GETGYIPSNVVA-PV-D 142  
Qy 55 LFLALYDFTARCAGGELSYYRRDRDLCALCEGGYIF-ARRLSGOPSAGLVPIITHAKASPE 113  
Db 143 SIQAEWYFGKLRKDAERQLLSFGNPRGTFTLIRESETTKGAYSLSIIRDWDMKGDHVKH 202  
Qy 55 LFLALYDFTARCAGGELSYYRRDRDLCALCEGGYIF-ARRLSGOPSAGLVPIITHAKASPE 113  
Db 143 SIQAEWYFGKLRKDAERQLLSFGNPRGTFTLIRESETTKGAYSLSIIRDWDMKGDHVKH 202

Qy 114 TLDSPQWFGSVRTOAQOALLSPNEPQAFILRSESSLGYSLSVRA-QA-K---VCH 168  
Db 203 YKIRLDNGGYITTRAQFETLQQLVQHYSEKADGCLFNLTVIASSCTPQTSGLAKDAWE 262  
Qy 169 YRVSMADGSLYLQKGR-LFPGLEELTY-KANWK-L-IQNPLQP-CMPQKAP-RQDVWE 224  
Db 263 VARRSLCLBKKGQGCFAEVWLTGNTGNTKVAIKTLKPGTMSPESELEEAQIMKKLKHDK 322  
Qy 225 RPHEFALGRKLGEYFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIQTLKGLRHER 284  
Db 323 LYQLYAVVSE-EPIYIVTEYMNKGSLLDLKDGEGRALKLPNLVDMAAOVAGMAYIERM 381  
Qy 285 LIRLHAVCSGGEPIVITELMRKGNLQAFLTGPEGRALRPLLLGFACQVAGMSYLEEQ 344  
Db 382 NYIHRDLRSANILVGNGLCKIADFLGARIIDSEYTAQEGAKFPKWTAPERALLYGRT 441  
Qy 345 RVVHRDLAARNVLVDGLACKVADFLGRLKDDIYSPSSSKIPVKTWTAPEAANYRVFS 404  
Db 442 IKSDVMSFGVLLMEVTVGRVPYPGNNREVLQVGRYMRPDTCPELYRGVTAECW 501  
Qy 405 QKSDVMSFGVLLHEVFTYGCQPYEGMTNHTLQQIMRGYRLPRPAACPAEVY-VLMLECW 464  
Db 502 KDPERPTEYLQ 514  
Qy 465 SSPEERPSFATLR 477

Search completed: Thu May 20 12:28:10 1999  
Job time : 327 secs.



\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:31:26 1999; MasPar time 8.25 Seconds  
Tabular output not generated.  
.....141.757-Million cell updates/sec

Title: >US-09-099-053-2  
Description: (1-55) from US09099053.pcp (2 of 6)  
Perfect Score: 429  
Sequence: 1 MEPTLRRRLAFLSFFWDKIW.....PNTDVPVTLPAEPCSPFPOL 55

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-genseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 26.823; Variance 102.975; scale 0.260

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	429	100.0	488	39	W89248	1.38e-34
2	85	19.8	563	18	R32128	8.95e-00
3	85	19.8	563	15	R79924	8.95e-00
4	82	19.1	302	35	W2007	1.60e-01
5	82	19.1	302	36	W2168	1.60e-01
6	82	19.1	302	27	W2120	1.60e-01
7	81	18.9	312	5	R26079	1.93e-01
8	81	18.9	340	37	W1594	1.93e-01
9	81	18.9	399	37	W1590	1.93e-01
10	79	18.4	249	35	W76166	2.82e-01
11	79	18.4	249	27	W31220	2.82e-01
12	79	18.4	249	24	W26460	2.82e-01
13	79	18.4	295	37	W68493	2.82e-01
14	79	18.4	400	32	W37431	2.82e-01
15	79	18.4	400	25	W23601	2.82e-01
16	79	18.4	400	25	W22521	2.82e-01

17	79	18.4	410	21	W12093	Bovine papillomavirus	2.82e-01
18	79	18.4	413	27	W31661	Bovine papillomavirus	2.82e-01
19	79	18.4	413	27	W31663	Bovine papillomavirus	2.82e-01
20	79	18.4	413	27	W31662	Bovine papillomavirus	2.82e-01
21	79	18.4	413	27	W31660	Bovine papillomavirus	2.82e-01
22	79	18.4	413	27	W31658	Bovine papillomavirus	2.82e-01
23	78	18.2	315	16	R48707	G-protein coupled hum	3.41e-01
24	78	18.2	315	19	W02679	G-protein coupled hum	3.41e-01
25	78	18.2	462	32	W57433	Cloned alkaline endog	3.41e-01
26	76	17.7	387	13	R70734	Human D4 dopamine rec	4.96e-01
27	76	17.7	387	21	W01749	Human dopamine D4 rec	4.96e-01
28	76	17.7	387	18	R96213	Recombinant human D4	4.96e-01
29	76	17.7	387	9	R48948	Sequence encoded by a	4.96e-01
30	76	17.7	387	5	R25335	D4 dopamine receptor	4.96e-01
31	76	17.7	387	14	R75957	Human dopamine D4 rec	4.96e-01
32	76	17.7	387	30	W40503	Human dopamine D4 rec	4.96e-01
33	76	17.7	419	9	R48949	Sequence encoded by a	4.96e-01
34	76	17.7	419	18	R96214	Recombinant human D4	4.96e-01
35	76	17.7	467	18	R96215	Recombinant human D4	4.96e-01
36	76	17.7	467	9	R48950	Sequence encoded by a	4.96e-01
37	75	17.5	235	13	R77286	Murine syndecan-1/hum	5.98e-01
38	75	17.5	348	1	P94683	Amino acid sequence e	5.98e-01
39	75	17.5	489	17	R88096	Thermostable beta-gal	5.98e-01
40	75	17.5	599	22	W09872	Rat huntingtin associ	5.98e-01
41	75	17.5	629	22	W09873	Rat huntingtin associ	5.98e-01
42	74	17.2	227	36	W72169	HSV-2 strain SB5 Cont	7.19e-01
43	74	17.2	758	31	W46270	Moraxella catarrhalis	7.19e-01
44	74	17.2	763	33	W55095	Streptococcus pneumon	7.19e-01
45	73	17.0	1958	12	R60620	Protein from ORF2 of	8.65e-01

## ALIGNMENTS

RESULT 1

ID W89248 standard; Protein; 488 AA.  
AC W89248;  
DT 10-MAR-1999 (first entry)  
DE Human SAD.  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease.  
OS Homo sapiens.  
PN W09849317-A2.  
PD 05-NOV-1998.  
PE 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGS) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plowman GD;  
DR WPI; 99-009434/01.  
DR N-PSDB; V81743.  
PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Claim 2; Page 154-155; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC present sequence represents human SAD. The above proteins, other than  
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
CC substances that modulate their activity (i.e. agonists and antagonists,  
CC including NBP) in vivo or in vitro. These substances are used to treat  
CC or prevent diseases associated with abnormal signal transduction  
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
CC serine/threonine kinase) are used to promote neuronal survival,  
CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
CC diseases. Nucleic acid fragments of the polynucleotides encoding the

CC proteins can be used as probes to identify and clone related sequences;  
 CC to detect protein-encoded RNA; to generate transgenic animals and in  
 CC gene therapy (optionally after mutation). Ab are used to determine the  
 CC proteins.  
 SQ Sequence 488 AA;

Query Match 100.0%; Score 429; DB 39; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 1.38e-34;  
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 mepflrrlafisffwdkiwpgagdhgtgsgldpntdpvptlpaeqpcspfqql 55  
 |||||  
 QY 1 MEPLFRLRLAFLSFFWDKIWPAGGEDHGTGSGLDPNTDPVPTLPAPCSPFPQL 55

# RESULT 2

ID R92128 standard; Protein; 563 AA.  
 AC R92128;  
 DT 17-OCT-1996 (first entry)  
 DE Porphyromonas gingivalis 72 kD subunit protein.  
 KW Periodontitis; gingivitis; type II fimbrial; vaccine; diagnosis;  
 OS Immunisation; epitope.  
 QS Porphyromonas gingivalis.  
 FH Key Location/Qualifiers  
 FT peptide 1..5  
 FT /label= sig\_peptide  
 FT protein 6..563  
 FT /label= mat\_protein  
 J08048695-A.  
 PD 20-FEB-1996.  
 PF 05-AUG-1994; 204422.  
 PR 05-AUG-1994; JP-204422.  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA (KYOW ) KYOWA MEDEX KK.  
 PA (MEIT ) MEITO SANGYO KK.  
 DR WPI: 96-167222/17.  
 DR N-PSDB; Tbl19.  
 PT Peptide(s) derived from type II fimbrial protein of Porphyromonas  
 PT gingivalis - used in the diagnosis and treatment of periodontitis  
 PS Example 1; Page 16-21; 22pp; Japanese.  
 CC R92128 is the 72 kD subunit protein of Porphyromonas gingivalis which  
 CC contains the type II fimbrial protein. Peptides derived from the  
 CC type II fimbrial protein are used in a compsn. for the diagnosis of  
 CC periodontitis. The peptides may also be used in a vaccine for  
 CC immunisation against the disease.  
 SQ Sequence 563 AA;

Query Match 19.8%; Score 85; DB 18; Length 563;  
 Best Local Similarity 31.8%; Pred. No. 8.95e-00;  
 Matches 14; Conservative 11; Mismatches 18; Indels 1; Gaps 1;

Db 497 lgfnwnplvpdpdpnsnpnnpdpnpgtgvptdpeglp 540  
 |||||  
 QY 12 LSFFWDKIWPAGGEDHGTGSGLDPNTD-PVPTLPAPCSPFPQ 54

# RESULT 3

ID R79924 standard; Protein; 563 AA:  
 AC R79924;  
 DT 10-MAY-1996 (first entry)  
 DE P. gingivalis cell surface protein.  
 KW Cell surface polypeptide; P. gingivalis; Bacteroides gingivalis;  
 KW PCSP409; diagnosis; prevention; periodontal disease; vaccine; ds.  
 OS Porphyromonas gingivalis strain OMZ409.  
 FH Key Location/Qualifiers  
 FT peptide 1..15  
 FT /note= "Signal peptide encoded by nucleotides designated  
 FT in the specification to encode the signal  
 FT peptide, comprises amino acids indicated as -5  
 FT to +10"  
 FT 50..563  
 FT /note= "Mature protein encoded by nucleotides designated  
 FT in the specification to encode the mature

FT peptide, comprises amino acids indicated as 45  
 FT to 558"  
 PN WO9526404-A1.  
 PD 05-OCT-1995.  
 PF 29-MAR-1995; J00584.  
 PR 29-MAR-1994; JP-081074.  
 PR 08-JUL-1994; JP-180815.  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA (KYOW ) KYOWA MEDEX CO LTD.  
 PA (MEIT ) MEITO SANGYO KK.  
 PI Fukui M, Hasegawa M, Hokkoku H, Mori H, Ogawa T;  
 PI Yamada K, Yasuda K;  
 DR WPI: 95-351324/45.  
 DR N-PSDB; T04121.  
 PT DNA coding for Porphyromonas gingivalis cell surface peptide - used  
 PT for production of the peptide for the prevention and diagnosis of  
 PT periodontal disease  
 PS Claim 1; Page 16-22; 30pp; Japanese.  
 CC This sequence represents the cell surface polypeptide of P. gingivalis.  
 CC The cDNA encoding this sequence was isolated on a 3.4 kb KpnI/XhoI P.  
 CC gingivalis fragment which was combined with paluescript to give pCSP409.  
 CC This plasmid was used to transform E. coli, which upon culture produced  
 CC the cell surface protein. The expressed protein can be used in the  
 CC diagnosis and prevention of periodontal disease, e.g. by  
 CC incorporation in a vaccine.  
 SQ Sequence 563 AA;

Query Match 19.8%; Score 85; DB 15; Length 563;  
 Best Local Similarity 31.8%; Pred. No. 8.95e+00;  
 Matches 14; Conservative 11; Mismatches 18; Indels 1; Gaps 1;

Db 497 lgfnwnplvpdpdpnsnpnnpdpnpgtgvptdpeglp 540  
 |||||  
 QY 12 LSFFWDKIWPAGGEDHGTGSGLDPNTD-PVPTLPAPCSPFPQ 54

# RESULT 4

ID W72007 standard; Protein; 302 AA.  
 AC W72007;  
 DT 07-DEC-1998 (first entry)  
 DE HSV-2 strain S85 Contig ID 101 ORF#1 protein.  
 KW HSV-2 strain S85; immunological response induction; therapy;  
 KW antiviral identification; viral protein inhibitor.  
 OS Herpes simplex virus type 2.  
 PN WO9820016-A1.  
 PD 14-MAY-1998.  
 PF 31-OCT-1997; U20016.  
 PR 09-JUN-1997; US-049018.  
 PR 04-NOV-1996; US-030279.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB,  
 PI Esser KM, Leary JJ;  
 DR WPI: 98-286847/25.  
 DR N-PSDB; V62131.  
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
 PT and treatment of infection or inducing immunological response in  
 PT mammal  
 PS Claim 10; Page 41; 748pp; English.  
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
 CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
 CC S85 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 101.  
 CC Based on homology, this sequence is a (X02138) 34K (Usl10) protein.  
 CC The proteins can be used for the treatment or prevention of disease, to  
 CC induce an immunological response in a mammal or to identify inhibitors,  
 CC activators or novel antivirals. Antagonists of the proteins can be used  
 CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
 CC it can also be used to induce an immunological response in a mammal.  
 SQ Sequence 302 AA;

Query Match 19.1%; Score 82; DB 35; Length 302;  
 Best Local Similarity 42.9%; Pred. No. 1.60e+01;  
 Matches 18; Conservative 9; Mismatches 9; Indels 6; Gaps 6;

CC The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors of activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal.

Query Match 19.1%; Score 82; DB 36; Length 302;  
Best Local Similarity 42.9%; Pred. No. 1.60e+01;  
Matches 18; Conservative 9; Mismatches 9; Indels 6; Gaps 6;

Db 65 raypgtrdphdpgcgsldphnpgaglp-p-vpyapl 104  
QY 18 KIWPAGGEP-D-HGTGSLDPTDPV-PT-LPAEPCSPFPQL 55

RESULT 7

ID R26079 standard; Protein; 312 AA.  
AC R26079;  
DT 03-FEB-1993 (first entry)  
DE Putative tat-E2 fusion protein.  
KW E2 protein; tat; HIV1; non-specific endocytosis; repressor.  
OS Synthetic.  
PN WO9212728-A.  
PD 06-AUG-1992.  
PF 28-JAN-1992; U00652.  
PR 28-JAN-1991; US-646998.  
PA (BIOI ) BIOGEN INC.  
PI (NEW-) NEW ENGLAND MEDICAL CENT HOSPITALS.  
PI Androphy EJ, Barsoum JG;  
WPI; 92-284418/34.  
DR N-PSDB; Q27279.  
CC E2 trans-activation repressors of Papilloma-virus - dimerise with full-length E2 polypeptide(s), for treating Papilloma-virus infections e.g. malignancies  
PS Disclosure; Page 92-4; 122pp; English.  
CC A preparation is described in the specification for the production of a tat-E2 fusion protein. The sequence given is the same length as the disclosed compound but it is not actually claimed as such. The fusion construct consisted of the first 62 amino acids of the tat protein from HIV1, a linker serine residue and the last 249 amino acids of BPV1 E2 protein. The fusion protein gained entry to cells via binding of the tat molecule with the cell surface and the non-specific endocytosis. The E2 molecule acts as a repressor once inside the cell.  
SQ Sequence 312 AA;

Query Match 18.9%; Score 81; DB 5; Length 312;  
Best Local Similarity 33.3%; Pred. No. 1.93e+01;  
Matches 14; Conservative 15; Mismatches 9; Indels 4; Gaps 4;

Db 112 dgywvasepe-gdpagkaeapqpvssllgspacqirpql 152  
QY 17 DKIWAGGEPDHGTGSLDPTDPV-PT-LPAEPCSPFPQL 55

RESULT 8

ID W81594 standard; Protein; 340 AA.  
AC W81594;  
DT 09-FEB-1999 (first entry)  
DE Protein encoded by human UCP3sh gene reading frame 3.  
KW Uncoupling protein 3; UCP3; thermogenesis; mammal; enhancer; drug;  
KW protein catabolism; anti-obesity; inhibitor; muscle wasting; infection;  
KW HIV; cancer; tumour cachexia; muscle disease; muscular dystrophy;  
KW non-insulin dependent diabetes mellitus; diagnosis; human.  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT 1..340  
FT /note= "the encoding reading frame has internal stop codons which are not indicated in this protein"

PD WO9845438-A1.  
PD 15-OCT-1998.



```

17 DKIWPGAGPDHGTGSLDPN-TDPVPTLPAP-CPSP 52
17 DKIWPGAGPDHGTGSLDPN-TDPVPTLPAP-CPSP 52

RESULT 11
ID W31220 standard; Peptide; 249 AA.
AC W31220;
DE 20-MAR-1998 (first entry)
DE HIV-TAT protein transport moiety peptide 11.
KW Human immunodeficiency virus; HIV Type 1; tat protein;
KW cargo molecule; intracellular delivery; fusion protein;
KW therapeutic; prophylactic; diagnostic; transport polypeptide;
KW E2 repressor protein.
OS Human immunodeficiency virus type 1.
PN US5674980-A.
PD 07-OCT-1997.
PF 21-DEC-1989; 454450.
PR 28-APR-1994; US-235403.
PR 21-DEC-1989; US-454450.
PR 02-JAN-1991; US-636662.
PR 21-AUG-1992; US-934375.
PR 19-AUG-1993; WO-007833.
PR 24-NOV-1993; US-158015.
PR 25-MAY-1995; US-450098.
PA (BARS/) BARSOUM J G.
PA (FAWE/) FAWELL S E.
PA (FRAN/) FRANKEL A.
PA (PABO/) PABO C.
PA (PEPI/) PEPINSKY R B.
PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
WPI: 97-502388/46.
PT Fusion proteins containing truncated HIV tat sequences - useful for
PT intracellular delivery of viral repressor proteins
PS Disclosure: Column 103-104; 77pp; English.
CC This peptide is used to devise a novel method for delivery of
CC biologically active cargo molecules into the cytoplasm and nuclei of
CC eukaryotic cells. The tat protein from immunodeficiency virus (e.g.
CC HIV-1, HIV-2, SIV) is readily taken up into cells when present
CC extracellularly and can be modified to covalently link to cargo proteins
CC e.g. E2 repressor proteins producing a fusion protein without the
CC problems of spurious trans-activation and disulphide aggregation. These
CC transport polypeptides also minimise interference with the biological
CC activity of the cargo molecule. This is applicable for therapeutic,
CC prophylactic or diagnostic intracellular delivery of small molecules and
CC macromolecules e.g. proteins, nucleic acids and polysaccharides.
SQ Sequence 249 AA;

Query Match 18.4%; Score 79; DB 27; Length 249;
Best Local Similarity 31.6%; Pred. No. 2.82e+01;
Matches 12; Conservative 15; Mismatches 8; Indels 3; Gaps 3;

Db 49 dqvwasegpe-gdpagkeapagpvssllgspacgpi 85
QY 17 DKIWPGAGPDHGTGSLDPN-TDPVPTLPAP-CPSP 52

RESULT 13
ID W68493 standard; Protein; 295 AA.
AC W68493;
DE 07-JAN-1999 (first entry)
DE A truncated papillomavirus E2 protein designated E2TR.
KW E2TR; papillomavirus; BVP-1; E2 protein; apoptosis; HPV; infection;
KW papillomavirus-associated cancer; cervix; virus-infected cell;
KW p53 tumour repressor.
OS Bovine papillomavirus.
PN WO9832861-A1.
PD 30-JUL-1998.
PF 29-JAN-1998; F00169.
PR 29-JAN-1997; FR-000964.
PA (INSP) INST PASTEUR.
PA (UYME-) UNIV MEXICO NACIONAL AUTONOMA.
PI Deneret C, Desaintes C, Goyat S, Thierry F, Yaniv M;
WPI: 98-427937/36.
DR N-PSDB: V60832.
DR Papilloma virus E2 protein or nucleic acid encoding it - useful for
PT treating and preventing cancer of the cervix
PS Disclosure: Fig 10; 111pp; French.
CC The present sequence represents E2TR, a truncated Bovine papillomavirus
CC type 1 (BPV-1) E2 protein. The E2TR protein does not contain the
CC N-terminal transactivation domain. Compounds derived from E2 are able
CC to induce apoptosis in cells that have integrated part of the human
CC papillomavirus (HPV) genome. The E2 protein and its derivatives, the
CC vectors (including those expressing wild-type E2) and the corresponding
CC proteins (including truncated E2 (E2TR)) are all useful for treating or preventing
CC papillomavirus infection, particularly papillomavirus-associated cancers
CC (especially of the cervix uteri). The proteins, and sequences expressing
CC them, also induce apoptosis of virus-infected cells and increasing the
CC activity of the p53 tumour repressor.
SQ Sequence 295 AA;

Query Match 18.4%; Score 79; DB 37; Length 295;
Best Local Similarity 31.6%; Pred. No. 2.82e+01;
Matches 12; Conservative 15; Mismatches 8; Indels 3; Gaps 3;

Db 78 dqvwasegpe-gdpagkeapagpvssllgspacgpi 114
QY 17 DKIWPGAGPDHGTGSLDPN-TDPVPTLPAP-CPSP 52
```



(TM)

Result No.	Query		Length	DB ID	Description	Pred. No.
	Match	Score				
1	231	53.8	496	2	156322	4.08e-26
2	231	53.8	496	2	A56040	4.08e-26
3	90	21.0	302	3	A36188	1.28e-01
4	87	20.3	633	2	B70946	3.38e-01
5	85	19.8	282	2	S74644	6.39e-01
6	83	19.3	488	2	A25156	1.20e+00
7	81	18.9	514	2	A35638	2.22e+00
8	80	18.6	1045	2	G69167	3.02e+00
9	79	18.4	306	1	W2WLEB	4.08e+00
10	79	18.4	325	2	S57089	4.08e+00
11	79	18.4	999	2	S68689	4.08e+00
12	78	18.2	365	2	S74847	4.08e+00
13	77	17.9	190	2	E64532	5.52e+00
14	77	17.9	257	2	H69756	7.43e+00
15	77	17.9	340	1	WMBEL1	7.43e+00
16	77	17.9	1008	2	S28858	7.43e+00
17	76	17.7	375	2	D70675	7.43e+00
18	76	17.7	387	1	DYHD04	9.99e+00
19	76	17.7	517	2	S77255	9.99e+00
20	75	17.5	316	2	A59021	1.34e+01
21	75	17.5	325	2	S02170	1.34e+01
22	75	17.5	385	2	S78100	1.34e+01
23	75	17.5	489	2	J00767	1.34e+01

```

#accession      targeted disruption.
##status        preliminary
##molecule_type mRNA
##residues      1-496 ##label KOH
##cross-references GB:D26186; NID:g529072; PID:d1005873; PID:g529073
GENETICS
#map_position 2
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
                kinases; protein kinase homology; SH3 homology
KEYWORDS      ATP; phosphotransferase
FEATURE
62-111        #domain SH3 homology #label SH3\
232-491        #domain protein kinase homology #label KIN\
240-248        #region protein kinase ATP-binding motif
SUMMARY      #length 496 #molecular-weight 55593 #checksum 301

Query Match      53.8%; Score 231; DB 2; Length 496;
Best Local Similarity 58.2%; Pred. No. 4.08e-26;
Matches 32; Conservative 7; Mismatches 12; Indels 4; Gaps 3;

Db 1 MEFPLRKRLTFLSFFWDKIWPAD-ESEEDIPRIQGHDDNPVPEQAAVPCS-PP 53
QY 1 MEFPLRRRLAFLSFFWDKIWPAGGEDHGTGSLDPTDVPVTLPA--EPCSPFP 53

RESULT 3
ENTRY   S36188      #type complete
TITLE   beta-lactamase (EC 3.5.2.6) I precursor - Streptomyces
        lactamdurans
ORGANISM #formal_name Streptomyces lactamdurans
DATE     13-Jan-1995 #sequence_revision 06-Jun-1997 #text_change
        08-Sep-1997
ACCESSION S36188; S22750
REFERENCE Coque, J.J.R.; Liras, P.; Martin, J.F.
EMBO J. (1993) 12:631-639
#journal
#title   Genes for a beta-lactamase, a penicillin-binding protein and
        a transmembrane protein are clustered with the cephamycin
        biosynthetic genes in Nocardia lactamdurans.
#accession S36188
#molecule_type DNA
#status    nucleic acid sequence not shown; translation not shown
#residues 1-302 ##label COQ
##cross-references EMBL:Z13971; NID:g44996; PID:g581412
##note    the nucleotide sequence was submitted to the EMBL Data
        Library, June 1992

GENETICS
#gene     bla
#start_codon GTG
CLASSIFICATION #superfamily beta-lactamase I
KEYWORDS      antibiotic resistance; hydrolase; penicillin resistance
FEATURE
1-29          #domain signal sequence #status predicted #label SIG\
30-302        #product beta-lactamase I #status predicted #label MAT\
85            #active_site Ser #status predicted
SUMMARY      #length 302 #molecular-weight 32084 #checksum 8442

Query Match      21.0%; Score 90; DB 2; Length 302;
Best Local Similarity 35.3%; Pred. No. 1.28e-01;
Matches 12; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 10 WARAPAAPAPPPSAAPSVAPGPAATPPDP 43
QY 16 WDKIPAGGPDHGTGSLDPTDVPVTL-PAEP 48

RESULT 4
ENTRY   B70946      #type complete
TITLE   probable nuoL protein - Mycobacterium tuberculosis (strain
        H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE     17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change

```

```

17-Jul-1998
ACCESSIONS      B70946
REFERENCE       A70500
#authors        Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
                III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
                Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                Fells, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                Hortsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
                Taylor, K.; Whitehead, S.; Barrrell, B.G.
#journal        Nature (1998) 393:537-544
#title          Deciphering the biology of Mycobacterium tuberculosis from
                the complete genome sequence.
#cross-references MUID:98295987
#accession      B70946
#status         preliminary; nucleic acid sequence not shown;
                translation not shown
#molecule_type DNA
#residues       1-633 ##label COL
##cross-references GB:AL021646; GB:AL123456; NID:g3242278; PID:el300744;
                PID:g3242279
##experimental_source strain H37Rv
GENETICS
#gene           nuoL
SUMMARY         #length 633 #molecular-weight 66167 #checksum 6554
                Query Match      20.3%; Score 87; DB 2; Length 633;
                Best Local Similarity 31.3%; Pred. No. 3.38e-01;
                Matches 10; Conservative 11; Mismatches 10; Indels 1; Gaps 1;

Db 434 AFYTRVMLMTFFGEKRWTPGAHP-HEAPAVM 464
QY 3 PFLLRRRLAFLSFFWDKIWPAGGPDHGTGSL 34

RESULT 5
ENTRY   S74644      #type complete
TITLE   regulatory component slr1783 - Synecocystis sp. (strain PCC
        6803)
ALTERNATE_NAMES hypotheetical protein slr1783
ORGANISM         #formal_name Synecocystis sp.
#variety         PCC 6803
DATE             25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                24-Sep-1998
ACCESSIONS      S74644
REFERENCE       S74322
#authors        Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
                Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                Yasuda, M.; Tabata, S.
#journal        DNA Res. (1996) 3:109-136
#title          Sequence analysis of the genome of the unicellular
                cyanobacterium Synecocystis sp. PCC6803. II. Sequence
                determination of the entire genome and assignment of
                potential protein-coding regions.
#cross-references MUID:97061201
#accession      S74644
#status         preliminary
#molecule_type DNA
#residues       1-282 ##label KAN
##cross-references EMBL:D90900; GB:AB001339; NID:g1651768; PID:d1017529;
                PID:g1651869
#note           the nucleotide sequence was submitted to the EMBL Data
                Library, June 1996
CLASSIFICATION #superfamily response regulator homology
FEATURE        58-170
                107
                #domain response regulator homology #label RRR\
                #binding_site phosphate (Asp) (covalent) #status

```



Db	450	GAEVDPPELPALPPQA-PLD-LFTQPPSPFHLH	491
Qy	23	GGE-PDHGTPGSLDPTDPVPTLPAEPCSPFQL	55
RESULT	8		
ENTRY	G69167	#type complete	
TITLE	cobalamin biosynthesis protein N - Methanobacterium thermoautotrophicum (strain Delta H)		
ORGANISM	#formal_name Methanobacterium thermoautotrophicum		
DATE	05-Dec-1997	#sequence_revision 05-Dec-1997	#text_change 05-Jun-1998
ACCESSIONS	G69167		
REFERENCE	A69000		
#authors	Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.		
#journal	J. Bacteriol.	(1997)	179:7135-7155
#title	Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.		
#cross-references	MUID:98037514		
#accession	G69167		
#status	preliminary; nucleic acid sequence not shown; translation not shown		
#molecule_type	DNA		
#residues	1-1045	#label MTH	
#cross-references	GB:AE000835; GB:AE000666; NID:g2621586; PID:g2621586		
#experimental_source	strain Delta H		
GENETICS			
#gene	MTH514		
SUMMARY	#length 1045	#molecular-weight 116537	#checksum 7198
Query Match	18.6%;	Score 80;	DB 2; Length 1045;
Best Local Similarity	30.6%;	Pred. No. 3.02e+00;	
Matches	11;	Conservative 10;	Mismatches 13; Indels 2; Gaps 2
Db	422	NPYLPPHHQYLAFYRWIDEVLGADAMVHLGTHGLE	457
Qy	2	EPFLRRRLAFLSFF-W-DKIWPAGPEPDHGTGSLD	35
RESULT	9		
ENTRY	W2WLEB	#type complete	
TITLE	E2 protein - bovine papillomavirus type 1		
ORGANISM	#formal_name bovine papillomavirus type 1		
DATE	17-Dec-1982	#sequence_revision 17-Dec-1982	#text_change 16-Feb-1997
ACCESSIONS	A03672		
REFERENCE	A93289		
#authors	Chen, E.Y.; Howley, P.M.; Levinson, A.D.; Seeburg, P.H.		
#journal	Nature (1982) 299:529-534		
#title	The primary structure and genetic organization of the bovine papillomavirus type 1 genome.		
#cross-references	MUID:83012974		
#accession	A03672		
#molecule_type	DNA		
#residues	1-306	#label CHE	
CLASSIFICATION	#superfamily papillomavirus E2 protein		
KEYWORDS	DNA binding; early protein; transcription regulation		
SUMMARY	#length 306	#molecular-weight 34307	#checksum 1656
Query Match	18.4%;	Score 79;	DB 1; Length 306;
Best Local Similarity	31.6%;	Pred. No. 4.08e+00;	
Matches	12;	Conservative 15;	Mismatches 8; Indels 3; Gaps 3

Db	210	DGWWASEGPE-GDPACKEAEPAQPVSSLLGSPACGP1-246	
QY	17	DKIWPAGGEPDHTGSLDPN-TDPVTLPAEP-CSFP 52	
RESULT	10	S57089 #type complete	
ENTRY		hypothetical protein YJ070c - yeast (Saccharomyces cerevisiae)	
TITLE		hypothetical protein J1814	
ALTERNATE_NAMES		#formal_name Saccharomyces cerevisiae	
ORGANISM		08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change	
DATE		05-Jun-1998	
ACCESSIONS		S57089; S71692	
REFERENCE		S57085	
#authors		Manus, V.; Huang, M.E.; Galibert, F.	
#submissions		submitted to the Protein Sequence Database, September 1995	
#accession		S57089	
#molecule_type		DNA	
#residues		1-325 #label MAN	
#cross-references		EMBL:Z49570; NID:g1015750; PID:g1015751; MIPS:YJ070c	
REFERENCE		S71676	
#authors		Huang, M.E.; Manus, V.; Chuat, J.C.; Galibert, F.	
#journal		Yeast (1996) 12:869-875	
#title		Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open reading frames and a gene cluster with a counterpart on chromosome XI.	
#accession		S71692	
#status		nucleic acid sequence not shown; translation not shown	
#molecule_type		DNA	
#residues		1-325 #label HUA	
#cross-references		EMBL:L47993; NID:g1019675; PID:g1019692	
#note		the nucleotide sequence was submitted to the EMBL Data Library, October 1995	
GENETICS			
#map_position		10R	
SUMMARY		#length 325 #molecular-weight 36164 #checksum 3014	
Query Match		18.4%; Score 79; DB 2; Length 325;	
Best Local Similarity		36.68; Pred. No. 4.08e+00;	
Matches		15; Conservative 7; Mismatches 17; Indels 2; Gaps 2;	
Db	151	NR1NTHGGAKRENQOOSYSSIDPAPPLPKDKATIPEL 191	
QY	17	DKI-WPAGGEPDHTGSLDPN-TDPVTLPAEP-CPSPQL 55	
RESULT	11	S68689 #type complete	
ENTRY		glucose regulated-protein, 170k--Chinese hamster	
TITLE		#formal_name Cricetulus griseus #common_name Chinese hamster	
ORGANISM		15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change	
DATE		10-Sep-1997	
ACCESSIONS		S68689	
REFERENCE		S68689	
#authors		Chen, X.; Easton, D.; Oh, H.J.; Lee-Yoon, D.S.; Liu, X.; Subject, J.	
#journal		FEBS Lett. (1996) 380:68-72	
#title		The 170 kDa glucose regulated stress protein is a large HSP70-, HSP110-like protein of the endoplasmic reticulum.	
#accession		S68689	
#status		preliminary; nucleic acid sequence not shown; translation not shown	
#molecule_type		mRNA	
#residues		1-999 #label CHE	
#cross-references		EMBL:U34206; NID:g1000295; PID:g1000296	
#note		the nucleotide sequence was submitted to the EMBL Data Library, August 1995	
SUMMARY		#length 999 #molecular-weight 111270 #checksum 7243	
Query Match		18.4%; Score 79; DB 2; Length 999;	
Best Local Similarity		28.1%; Pred. No. 4.08e+00;	
Matches		9; Conservative 13; Mismatches 9; Indels 1; Gaps 1;	

```
##molecule_type DNA
##residues 1-190 #label TOM
##cross-references GB:AE000524; GB:AE000511; NID:g2313090; PID:g2313099;
TIGR:HP0021
SUMMARY #length 190 #molecular-weight 21183 #checksum 5365

Query Match 17.9%; Score 77; DB 2; Length 190;
Best Local Similarity 50.08; Pred. No. 7.43e+00;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 134 AFLMRRYSFKKFEW 147
QY 3 PFLRRRLAFLSFEW 16
: || || : ||
: || || : ||

RESULT 14
ENTRY #type complete
TITLE tellurium resistance protein homolog yceF - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS H69756
REFERENCE H69756
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haelech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashnara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
#journal Nature
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession H69756
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-257 #label KUN
##cross-references GB:299105; GB:AL009126; NID:g2632457; PID:el182244;
PID:g2632578
#experimental_source strain 168
GENETICS
#gene yceF

CLASSIFICATION #superfamily hypothetical protein sll1022
SUMMARY #length 257 #molecular-weight 29167 #checksum 3438

Query Match 17.9%; Score 77; DB 2; Length 257;
Best Local Similarity 45.5%; Pred. No. 7.43e+00;
Matches 10; Conservative 6; Mismatches 2; Indels 4; Gaps 4;

Db 1 MD-FLHILSTYASFFDW-KMW 20
QY 1 MEPLRRRLA-FUSFF-WDKIW 20
: || || : || || : ||
: || || : || || : ||

RESULT 15
ENTRY #type complete
TITLE latency-related protein 1 - human herpesvirus 1 (strain F)
ORGANISM #formal_name human herpesvirus 1
#note host Homo sapiens (man)
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
ACCESSIONS A33337
REFERENCE A94388
#authors Wechsler, S.L.; Nesburn, A.B.; Zwaagstra, J.; Ghiasi, H.
#journal Virology (1989) 168:168-172
#title Sequence of the latency-related gene of herpes simplex virus
type 1.
#cross-references MUID:89085598
#accession A33337
#status translation not shown
##molecule_type mRNA
#residues 1-340 #label WEC
##cross-references GB:J04323; NID:g330133; PID:g330134
GENETICS
#introns 249/2
CLASSIFICATION #superfamily herpesvirus latency-related protein 1
KEYWORDS tandem repeat
FEATURE
27-43,59-75 #region tandem repeats
SUMMARY #length 340 #molecular-weight 35604 #checksum 4407

Query Match 17.9%; Score 77; DB 1; Length 340;
Best Local Similarity 30.6%; Pred. No. 7.43e+00;
Matches 11; Conservative 11; Mismatches 13; Indels 1; Gaps 1;

Db 15 LMLPPEPAQHGTPPTTHPSHA-PPLPRTPTPSHPH 49
QY 19 IWPAGEPDHGTPGSLDPTDPTLPAEPCSPFPQ 54
: | : || || : | : || | : |
: | : || || : | : || | : |

Search completed: Thu May 20 12:35:50 1999
Job time : 112 secs.
```

\*\*\*\*\*

W P I S R E H (TW)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:36:09 1999; Maspar time 4.00 Seconds  
Tabular output not generated. 388.629 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (1-55) from US09099053.pep (2 of 6)  
Perfect Score: 429  
Sequence: 1 MEPTLRRRLAFLSFFWDKIW.....PNTDPVPTLPAPCSPPFPL 55

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 36.911; Variance 63.964; scale 0.577

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	231	53.8	496	1 SRM_MOUSE	TYROSINE-PROTEIN KINAS	2.66e+29
2	90	21.0	302	1 BLAC_NOCLA	BETA-LACTAMASE PRECURS	3.87e+02
3	83	19.3	488	1 GUN1_BACSA	ENDOGLUCANASE A (EC 3.	4.46e+01
4	81	18.9	514	1 TFEB_HUMAN	TFEB PROTEIN (FRAGMENT	8.77e+01
5	79	18.4	325	1 YU40_YEAST	HYPOTHETICAL 36.2 KD P	1.70e+00
6	78	18.4	410	1 VE2_BPV1	REGULATORY PROTEIN E2.	1.70e+00
7	78	18.2	324	1 FKH5_MOUSE	TRANSCRIPTION FACTOR F	2.36e+00
8	77	17.9	340	1 LRPI_HSV1F	LATENCY-RELATED PROTEI	3.27e+00
9	76	17.7	467	1 DADR_HUMAN	D(4) DOPAMINE RECEPTO	4.51e+00
10	76	17.7	1197	1 Y4CA_RHISN	HYPOTHETICAL 133.7 KD	4.51e+00
11	75	17.5	283	1 TONB_NEIGO	TONB PROTEIN.	6.20e+00
12	75	17.5	325	1 CA19_RAT	COLLAGEN ALPHA 1(IX) C	6.20e+00
13	75	17.5	385	1 MKK2_MOUSE	MAP KINASE-ACTIVATED P	6.20e+00
14	75	17.5	489	1 BGAL_SULSO	BETA-GALACTOSIDASE (EC	6.20e+00
15	75	17.5	489	1 BGAL_SULSH	BETA-GALACTOSIDASE (EC	6.20e+00
16	75	17.5	629	1 HAP1_RAT	HUNTINGTIN ASSOCIATED	6.20e+00
17	75	17.5	986	1 GNFI_DRONE	GERMLINE TRANSCRIPTION	6.20e+00
18	74	17.2	278	1 TONB_NEIME	TONB PROTEIN.	8.50e+00
19	74	17.2	543	1 VP61_NPVAC	61 KD PROTEIN.	8.50e+00
20	74	17.2	700	1 NUOL_RHOCA	NADH DEHYDROGENASE I C	8.50e+00
21	74	17.2	718	1 AT12_HSV11	ALPHA TRANS-INDUCING F	8.50e+00
22	74	17.2	1733	1 VNVA_PRVKA	PROBABLE NUCLEAR ANTIG	8.50e+00
23	74	17.2	2124	1 PGCA_RAT	AGGREGAN CORE PROTEIN	8.50e+00

24	73	17.0	219	1 ERD2_YEAST	ER LUMEN PROTEIN RETAI	1.16e+01
25	73	17.0	402	1 LIM2_MOUSE	HOMEOBOX PROTEIN LIM-2	1.16e+01
26	73	17.0	478	1 KCCA_RAT	CALCIUM/CALMODULIN-DEP	1.16e+01
27	73	17.0	478	1 KCCA_MOUSE	CALCIUM/CALMODULIN-DEP	1.16e+01
28	73	17.0	613	1 XPCT_HUMAN	X-LINKED PEST-CONTAINI	1.16e+01
29	73	17.0	802	1 Y250_HUMAN	HYPOTHETICAL PROTEIN K	1.16e+01
30	72	16.8	315	1 ALDR_PIG	ALDOSE REDUCTASE (EC 1	1.58e+01
31	72	16.8	326	1 CC14_CAEEL	CUTICLE COLLAGEN 14.	1.58e+01
32	72	16.8	424	1 THIK_RAT	3-KETOACYL-COA THIOLAS	1.58e+01
33	72	16.8	477	1 ETVI_MOUSE	BETA1 PROTEIN (ETS TRAN	1.58e+01
34	72	16.8	491	1 BGAL_SULAC	ERB1-GALACTOSIDASE (EC	1.58e+01
35	72	16.8	1595	1 SOS_DROME	SON OF SEVENLESS PROTE	1.58e+01
36	72	16.8	2132	1 PGCA_MOUSE	AGGREGAN CORE PROTEIN	1.58e+01
37	71	16.6	257	1 YXP2_YANCP	HYPOTHETICAL 26.9 KD P	2.14e+01
38	71	16.6	336	1 TER4_HUMAN	TRANSCRIPTIONAL ENHANC	2.14e+01
39	71	16.6	399	1 FTSZ_STRCO	CELL DIVISION PROTEIN	2.14e+01
40	71	16.6	435	1 YNJE_ECOLI	PUTATIVE THIOSULFATE S	2.14e+01
41	71	16.6	522	1 KAPR_USTMA	CAMP-DEPENDENT PROTEIN	2.14e+01
42	71	16.6	605	1 PHAC_METEX	POLY(3-HYDROXYALKANOAT	2.14e+01
43	71	16.6	634	1 HWP1_CANAL	HYPHAL WALL PROTEIN 1	2.14e+01
44	71	16.6	656	1 DNAA_STRCO	CHROMOSOMAL REPLICATIO	2.14e+01
45	71	16.6	1027	1 CAFF_RIFPA	FIBRIL-FORMING COLLAG	2.14e+01

ALIGNMENTS

RESULT 1	SRM_MOUSE	STANDARD:	PRT:	496 AA.
AC	Q62270; Q62360;			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	TYROSINE-PROTEIN KINASE SRM (EC 2.7.1.112) (PTK70).			
GN	SRMS OR SRM.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
CC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-C57BL/6; TISSUE-THYMUS;			
RX	MEDLINE; 97369678.			
RA	KAWACHI Y., NAKAUCHI H., OTSUKA F.;			
RT	"Isolation of a cDNA encoding a tyrosine kinase expressed in murine			
RT	skin.";			
RL	EXP. DERMATOL. 21:533-538(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LUNG;			
RX	MEDLINE; 95021220.			
RA	KOHMURA N., YAGI T., TOMOOKA Y., OYANAGI M., KOMINAMI R., TAKEDA N.,			
RA	CHIBA J., IKAWA Y., AIZAWA S.;			
RT	"A novel nonreceptor tyrosine kinase, Src: cloning and targeted			
RT	disruption.";			
RL	MOL. CELL. BIOL. 14:6915-6925(1994).			
CC	MO. CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +			
CC	PROTEIN TYROSINE PHOSPHATE.			
CC	-!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC			
CC	DOMAIN. BELONGS TO THE SRC SUBFAMILY.			
CC	-!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.			
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch)			
CC	-----			
DR	EMBL; D49427; G584972; -			
DR	EMBL; D26186; G529073; -			
DR	MGD; MGI:101865; SRMS.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.			

```

FT BINDING 247 249 SUBSTRATE (BY SIMILARITY).
SQ SEQUENCE 302 AA; 32084 MW; 3AF376AC CRC32;

Query Match 21.0%; Score 90; DB 1; Length 302;
Best Local Similarity 35.3%; Pred. No. 3.87e-02;
Matches 12; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 10 WARARPAAPPEPAPPTPSAAAPSVAPGPAATPPDP 43
   | : ||| : || ||| : | : | : | : |
QY 16 WDXIWPAAGPEPDHGTGSLDPNTDPVPTL-PAEP 48

RESULT 3
ID GUNL_BACS4 STANDARD; PRT; 488 AA.
AC P06366;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE)
DE (CLONE PNK1).
DE CELA.
OS BACILLUS SP. (STRAIN N-4).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 87056924.
RA FUKUMORI F., SASHIHARA N., KUDO T., HORIKOSHI K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RT Bacillus sp. strain N-4 and their strong homology.";
RL J. BACTERIOL. 168:479-485(1986).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14781; G142660; -
DR PIR; A25156; A25156.
DR PROSITE; P500659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
DR KJ; K01500; cellulase; 1.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE.
FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 488 AA; 54264 MW; 1C14BCFB CRC32;

Query Match 19.3%; Score 83; DB 1; Length 488;
Best Local Similarity 38.2%; Pred. No. 4.46e-01;
Matches 13; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

Db 414 PSEPEDDPGEPPDGEPPDGEPPDTPPSDP-GEYP 446
   | : ||| | : | : | : | : | : |
QY 21 PAGGEPDHTGPGSLDPNT-DPVPTLPAEPCSPPP 53

RESULT 4
ID TFEB_HUMAN STANDARD; PRT; 514 AA.
AC P19484;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TFEB PROTEIN (FRAGMENT).
DE HOMO SAPIENS (HUMAN).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]

```



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	ENBL; X02346; -; NOT_ANNOTATED_CDS.
DR	PIR; A03672; WZWLBB.
DR	PDB; 2BOP; 31-JAN-94.
DR	PFAM; PF00508; E2-N; 1.
DR	TRANSFAC; T00205; -.
KW	EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;
KW	TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN;
KW	PHOSPHORYLATION; 3D-STRUCTURE.
SW	SEQUENCE 410 AA; 45449 MW; FLD75660 CRC32;

Query Match 18.4%; Score 79; DB 1; Length 410;  
Best Local Similarity 31.6%; Pred. No. 1.70e+00;  
Matches 12; Conservative 15; Mismatches 8; Indels 3; Gaps 3;

Db 210 DGWVASGEPE-GDPAGKEAPEAQPVSSLLGSPACGP I 246  
| : | : | : | : | : | : | : | : | : | :  
Qy 17 DKIWPAGEPDHGTGSLDPN-TDVPPTLPAEP-CSPF 52

RESULT	7	FKH5_MOUSE	STANDARD;	PRT;	324 AA.
ID		Q64732;			
AC					
DT		01-NOV-1997	(REL. 35, CREATED)		
DT		01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)		
DT		01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)		
DE		TRANSCRIPTION FACTOR FKH-5.			
GN		FKH5 OR MF3.			
OS		MUS MUSCULUS (MOUSE).			
OC		EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC		RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.			

RP SEQUENCE FROM N.A.  
RA KAESTNER K.H., SCHUETZ G., MONAGHAN A.P.;  
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
[2]  
RP SEQUENCE OF 4-114 FROM N.A.  
RC STRAIN=129;  
RX MEDLINE: 93361500.  
RA KAESTNER K.H., LEE K.H., SCHLOENDORFF J., HIEMISCH H., MONAGHAN A.P.,  
RA SCHUETZ G.;  
RT "Six members of the mouse forkhead gene family are developmentally  
RL regulated.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 90:7628-7631(1993).

RP SEQUENCE OF 1-68 FROM N. A.  
RX MEDLINE; 9413970.  
RA ANG S.-L., WIERDA A., WONG D., STEVENS K.A., CASCIO S., ROSSANT J.,  
RZ ZARET K.S.;  
RT "The formation and maintenance of the definitive endoderm lineage in  
RL the mouse: involvement of HNF3/forhead proteins.";  
RL DEVELOPMENT 119:1301-1315(1993).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; X92592; E208624; -.  
DR EMBL; X71943; G311745; -.  
DR EMBL; U04198; G433397; -.  
DR MGD; MGI:103153; MF3.  
DR PROSITE; PS00657; FORK\_HEAD\_1; 1.

DR	PROSITE; PS00658;	FORK_HEAD_2;	1.
DR	PROSITE; PS50039;	FORK_HEAD_3;	1.
DR	PFAM; PF00250;	Fork_head;	1.
DR	TRANSFAC; T02443;	-	
DR	DNA-BINDING; NUCLEAR PROTEIN;	TRANSCRIPTION REGULATION.	
KW	DNA-BIND	12	103
FT	CONFLICT	63	68
FT	SEQUENCE	324 AA;	34937 MW;
SQ	HLSNFSN ->	PLTVRH (IN REF. 3).	
SC	30FACE2A	CRC32;	

```

Query Match      18.28; Score 78; DB 1; Length 324;
Best Local Similarity 36.23; Pred. No. 2.36e+00;
Matches 13; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

Db      250 AAGGTLPAPVPKPTPAAVPALPALP-APITPL 282
      : : : : : : : : : : : : : : : : : : :
Qv      22 AGGEPDHTGTSGLDPNDTPVTPLPAEPCSPFPL 55

```

RESULT 8  
ID LRP1\_HSV1F STANDARD; PRT; 340 AA.

DT 01-AUG-1990 (REL. 15, CREATED)  
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
DE LATENCY-RELATED PROTEIN 1.  
DS HERPES SIMPLEX VIRUS (TYPE 1 / STRAIN F).  
OC VIRUSSES; DSDNA VIRUSES, NO RNA STAGE; HERPESV  
OC ALPHAHERPESVIRINAE; SIMPLEXVIRUS.

AN . [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 8908598.  
RA WECHSLER S.L., ZWAGSTRA J., GHIASI H.;  
RX "Sequence of the latency-related gene of herpes simple  
RT 1.";  
RT  
RL VIROLOGY 168:168-172(1989).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

[illegible]

Query Match	17.9%	Score 77	DB 1	Length 340
Best Local Similarity	30.6%	Pred. No.	3.27e+00	
Matches	11	Conservative	11	Mismatches 13
		Indels	1	Gaps 1

```
Db 15 LWLTPEPAQHGTPtPHPSHA -PPLPRPTPSHPH 49
    :| : :||| : | : | || | :| :
Qv 19 IWPAGEPDHGTPGSLDNDTDPVTLPAEPCSPFO 54
```

RESULT	9	
ID	D4DR_HUMAN	STANDARD;
		PRT: 467 AA.

DT 01-MAY-1991 (REL. 18, CREATED)  
DT 01-FEB-1986 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
D(4) DOPAMINE RECEPTOR (D(2C) DOPAMINE RECEPTOR).  
DRD4.  
DN GN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]





```

15-DEC-1998 (REL. 37, CREATED)
15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TONB PROTEIN.
GN TONB.
OS NEISSERIA GONORRHOAE.
CC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIAEAE; NEISSERIA.
RN [1]
RN SEQUENCE FROM N.A.
RN STRAIN=FA19.
RN MEDLINE; 97285757.
RX BISWAS G.D., ANDERSON J.E., SPARLING P.F.;
RT "Cloning and functional characterization of Neisseria gonorrhoeae
RL tonB, exbB and exbD genes.";
RL MOL. MICROBIOL. 24:169-179(1997).
CC -1- FUNCTION: PATHWAYS OF UTILIZATION OF IRON BOUND TO TRANSFERRIN,
CC LACTOFERRIN AND HEMOGLOBIN BUT NOT TO HAEMIN OR CITRATE WHERE
CC DEPENDENT ON THE TONB SYSTEM.
CC -1- SUBUNIT: THE ACCESSORY PROTEINS EXBB AND EXBD SEEM TO FORM A
CC COMPLEX WITH TONB (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TONB FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U79563; G2098621.
CC TRANSPORT; PROTEIN TRANSPORT; INNER MEMBRANE; PERIPLASMIC;
CC TRANSMEMBRANE.
KW DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 6 27 ANCHOR (POTENTIAL).
FT DOMAIN 28 283 PERIPLASMIC (POTENTIAL).
FT SEQUENCE 283 AA; 28749 MW; 964F0942 CRC32;
-----
Query Match 17.5%; Score 75; DB 1; Length 283;
Best Local Similarity 38.7%; Pred. No. 6.20e+00;
Matches 12; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

Db 52 GGGAGGAGAGAAPEQPAPDPDPKPVEPPKP 82
QY 22 AGGEPD-HGTPGSLDPNTDPVPTLPAPFCSP 51
-----
RESULT 12
ID CA19.RAT STANDARD; PRT; 325 AA.
AC P20850;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE COLLAGEN ALPHA 1(IX) CHAIN (FRAGMENT).
GN COL2A1.
OS RATTUS NORVEGICUS (RAT).
CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RN SEQUENCE FROM N.A.
RN MEDLINE; 89137096.
RX KIMURA T., MATTEI M.-G., STEVENS J.W., GOLDRING M.B., NINOMIYA Y.,
RA OLSEN B.R.;
RT "Molecular cloning of rat and human type IX collagen cDNA and
RT localization of the alpha 1(IX) gene on the human chromosome 6.";
RL EUR. J. BIOCHEM. 179:71-78(1989).
CC -1- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
CC COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(IX),
CC ALPHA 2(IX), AND ALPHA 3(IX).

```

KW TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;  
KW PHOSPHORYLATION.

FT NON\_TER 1 1 PRO-RICH.  
FT DOMAIN 5 29 POLY-PRO.  
FT DOMAIN 12 21 POLY-PRO.  
FT DOMAIN 24 29 POLY-PRO.  
FT DOMAIN 49 310 PROTEIN KINASE.  
FT NP\_BIND 55 63 ATP (BY SIMILARITY).  
FT BINDING 78 78 ATP (BY SIMILARITY).  
FT ACT\_SITE 171 171 BY SIMILARITY.  
FT MOD\_RES 319 319 PHOSPHORYLATION (BY MAPK) (BY  
ET SIMILARITY).  
SQ SEQUENCE 385 AA; 43952 MW; E3629FEA CRC32;

Query Match 17.5%; Score 75; DB 1; Length 385;  
Best Local Similarity 40.6%; Pred. No. 6.20e+00;  
Matches 13; Conservative 9; Mismatches 7; Indels 3; Gaps 3;

Db 3 GSPGQTPAPF-PSPPP-PA-PQAPPPPPFQF 31

QY 24 GPDHGTGSLDPTDPTLPAEPCSPFQQL 55

RESULT: 14

ID BGAL-SULSO STANDARD; PRT; 489 AA.  
AC P22498;  
DT 01-AUG-1991 (REL. 19, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).  
GN LACS.  
OS SULFOLOBUS SOLFATARICUS.  
OC ARCHAEA; CRENARCHAEOTA; SULFOLOBALES; SULFOLOBUS.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-33.  
RC STRAIN-ATCC 49255 / MT4;  
RX MEDLINE; 91033070.  
RA CUBELLIS M.V., ROZZO C., MONTECUCCHI P., ROSSI M.;  
RT "Isolation and sequencing of a new beta-galactosidase-encoding  
archaeobacterial gene";  
RL GENE 94:89-94(1990).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
RC STRAIN-ATCC 49255 / MT4;  
RX MEDLINE; 97446327.  
RA AGUILAR C.F., SANDERSON I., MORACCI M., CIARAMELLA M., NUCCI R.,  
RA ROSSI M., PEARL L.H.;  
RT "Crystal structure of the beta-glycosidase from the hyperthermophilic  
archaeon Sulfolobus solfataricus: resilience as a key factor in  
thermostability";  
RL J. MOL. BIOL. 271:789-802(1997).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; M34696; G152929;  
DR PIR; JQ0767; JQ0767.  
DR PDB; 1GOW; 20-AUG-97.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_FL\_1; 1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_FL\_2; 1.  
DR PFAM; PF00232; glycosyl\_hydrol; 1.  
KW HYDROLASE; GLYCOSIDASE; 3D-STRUCTURE.  
FT ACT\_SITE 206 206 PROTON DONOR (POTENTIAL).  
FT ACT\_SITE 387 387 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 489 AA; 56757 MW; 6B8E9647 CRC32;

Query Match 17.5%; Score 75; DB 1; Length 489;  
Best Local Similarity 57.1%; Pred. No. 6.20e+00;  
Matches 12; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

Db 12 WSOAGFQSEMGTGSGEDPNTD 32

QY 20 WPAGG-EPDHGTGSLDPTD 39

RESULT: 15

ID BGAL-SULSH STANDARD; PRT; 489 AA.  
AC P30388;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).  
GN BGLY.  
OS SULFOLOBUS SHIBATAE.  
OC ARCHAEA; CRENARCHAEOTA; SULFOLOBALES; SULFOLOBUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DSM 5389;  
RA CONNARIS H., CHARALAMBOUS B.M.;  
RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; L47841; G1009227;  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_FL\_1; 1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_FL\_2; 1.  
DR PFAM; PF00232; glycosyl\_hydrol; 1.  
DR HSP; P22498; 1GOW.  
KW HYDROLASE; GLYCOSIDASE.  
FT ACT\_SITE 206 206 PROTON DONOR (POTENTIAL).  
FT ACT\_SITE 387 387 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 489 AA; 56564 MW; 06E7FBF5 CRC32;

Query Match 17.5%; Score 75; DB 1; Length 489;  
Best Local Similarity 57.1%; Pred. No. 6.20e+00;  
Matches 12; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

Db 12 WSOAGFQSEMGTGSGEDPNTD 32

QY 20 WPAGG-EPDHGTGSLDPTD 39

Search completed: Thu May 20 12:36:22 1999  
Job time : 13 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	91	21-2	2715	5	O61503	EYELID.	2.10e-01
2	89	20.7	4123	4	O73551	WUGSC:H_DJ0751H13.1 PR	3.96e-01
3	88	20.5	553	2	O83017	67KDA FIMBRILLIN.	5.43e-01
4	88	20.5	553	2	O24773	PGA67 PRECURSOR.	5.43e-01
5	88	20.5	553	2	O51819	CELL SURFACE PROTEIN P	5.43e-01
6	87	20.3	633	2	O86350	NUOL.	7.42e-01
7	86	20.0	770	5	O20942	SIMILAR TO S. PURPURAT	1.01e+00
8	85	19.8	282	2	F72781	REGULATORY COMPONENTS	1.38e+00
9	85	19.8	553	2	O51820	CELL SURFACE PROTEIN P	1.38e+00
10	83	19.3	498	14	P88150	GAG POLYPROTEIN.	2.54e+00
11	82	19.1	302	14	P89478	VIRION PROTEIN.	3.44e+00
12	82	19.1	408	5	O44776	F33D11.2 PROTEIN.	3.44e+00
13	82	19.1	1089	5	O26155	V-SERA 1.	3.44e+00
14	81	18.9	696	2	O54201	PBF2 PROTEIN.	4.64e+00
15	80	18.6	205	14	O84959	ORF 3-1 GENE.	6.24e+00
16	80	18.6	429	5	O44497	F15E6.2 PROTEIN.	6.24e+00
17	80	18.6	599	5	O18756	PROLINE AND GLYCINE-RI	6.24e+00
18	80	18.6	655	5	O15979	E6-AP UBIQUITIN-PROTEI	6.24e+00
19	80	18.6	705	2	O85369	HYPOTHETICAL 77.1 KD P	6.24e+00
20	80	18.6	1045	1	O26614	COBALAMIN BIOSYNTHESIS	6.24e+00

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;

```
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA LEONARD S., GRAVES T., STROMMATT C.;
RT "The sequence of Homo sapiens PAC clone DJ0751H13.";
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.H.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC004877; G3638957; -
DR PROSITE; PS01209; LDLRA_1; 9.
KW GLYCOPROTEIN.
FT NON_TER
SQ SEQUENCE 4123 AA; 434981 MW; 0D93C3B8 CRC32;

Query Match 20.7%; Score 89; DB 4; Length 4123;
Best Local Similarity 45.5%; Pred. No. 3.96e-01;
Matches 15; Conservative 7; Mismatches 9; Indels 2; Gaps 2;

Db 1333 PASTLPCP-SPGSLDTASSPLASAPCGPF 1364
||: | :||| | :|: | |||
QY 21 PAGGEDHGTGSLDPTDVPVTL-PAEPCSPF 52

RESULT 3 PRELIMINARY; PRT; 563 AA;
ID O83017
AC O83017;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE 67KDA FIMBRILLIN.
GN MFAL.
OS PORPHYROMONAS GINGIVALIS (BACTEROIDES GINGIVALIS).
OC BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33277;
RA HAMADA N., UMEMOTO T.;
RT "Porphyromonas gingivalis 67kDa fimbriillin gene.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB016284; D1032915; -
SQ SEQUENCE 563 AA; 60786 MW; 524635A6 CRC32;

Query Match 20.5%; Score 88; DB 2; Length 563;
Best Local Similarity 31.8%; Pred. No. 5.43e-01;
Matches 14; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

Db 497 LGFNWNPVLPDPPSNPENPNNDPDPGTPVPTDPENPLPD 540
||: | :| | :| :| | | :|: | |||
QY 12 LSFWDKIWPAGGPDHGTGSLDPTD-PVPTLPAEPCSPFPQ 54

RESULT 4 PRELIMINARY; PRT; 563 AA.
ID O24773
AC O24773;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMREL. 05, LAST ANNOTATION UPDATE)
DE PGA67 PRECURSOR.
GN PGA67.
OS PORPHYROMONAS GINGIVALIS (BACTEROIDES GINGIVALIS).
OC BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC33277;
RA HONGYO H., KOREGUCHI S., MAEDA H., MIYAMOTO M., TAKASHIBA S.,
RA KURIHARA H., MURAYAMA Y.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AB006481; D1023712; -
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 49 POTENTIAL.
SQ SEQUENCE 563 AA; 60811 MW; 204E6E96 CRC32;

Query Match 20.5%; Score 88; DB 2; Length 563;
Best Local Similarity 31.8%; Pred. No. 5.43e-01;
Matches 14; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

Db 497 LGFNWNPVLPDPPSNPENPNNDPDPGTPVPTDPENPLPD 540
||: | :| | :| :| | | :|: | |||
QY 12 LSFWDKIWPAGGPDHGTGSLDPTD-PVPTLPAEPCSPFPQ 54

RESULT 5 PRELIMINARY; PRT; 563 AA.
ID Q51819
AC Q51819;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE CELL SURFACE PROTEIN PRECURSOR.
OS PORPHYROMONAS GINGIVALIS (BACTEROIDES GINGIVALIS).
OC BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=381;
RA OGAWA T., MORI H., YASUDA K., HASEGAWA M.;
RT "Molecular cloning and characterization of the genes encoding the
RT immunoreactive major cell-surface proteins of Porphyromonas
RT gingivalis.";
RL FEMS MICROBIOL. LETT. 120:23-30(1994).
DR EMBL; D28770; D1006510; -
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 15 POTENTIAL.
SQ SEQUENCE 563 AA; 60785 MW; 45D4ECB2 CRC32;

Query Match 20.5%; Score 88; DB 2; Length 563;
Best Local Similarity 31.8%; Pred. No. 5.43e-01;
Matches 14; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

Db 497 LGFNWNPVLPDPPSNPENPNNDPDPGTPVPTDPENPLPD 540
||: | :| | :| :| | | :|: | |||
QY 12 LSFWDKIWPAGGPDHGTGSLDPTD-PVPTLPAEPCSPFPQ 54

RESULT 6 PRELIMINARY; PRT; 633 AA.
ID O86350
AC O86350;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE NUOL.
GN NUOL.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE.
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY III C.E., TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTLER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
```

```

DE REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEM.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUKI K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT genome and assignment of potential protein-coding regions.";
RL DNA RES. 3:109-136(1996).
DR EMBL; D90900; D1017529; -.
DR PFAM; PF00072; response_reg; 1.
DR PFAM; PF00196; Gefe; 1.
SQ SEQUENCE 282 AA; 31395 MW; BB8E81A9 CRC32;

Query Match 19.8%; Score 85; DB 2; Length 282;
Best Local Similarity 37.5%; Pred.No. 1.38e+00;
Matches 18; Conservative 9; Mismatches 17; Indels 4; Gaps 4;

Db 4 SLRLRLVLEFIPRDRL-MGGIGISQGNLLR-SPPTKP-PYLPLHDC 48
:| | | | | | | | | | | | | | | | | | | | |
Qy 3 PFLRRRLAELSEF-WDKINPAGGPDHGTGSLDPTNPVPTLPAEPC 49

RESULT 9
ID Q51820 PRELIMINARY; PRT; 563 AA.
AC Q51820;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CELL SURFACE PROTEIN PRECURSOR.
OC PORPHYROMONAS GINGIVALIS (BACTEROIDES GINGIVALIS).
OC BACTERIA; CYTOPHAGALES; BACTEROIDACEAE; PORPHYROMONAS.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=409;
RX MEDLINE; 94333763.
RA OGAWA T., MORI H., YASUDA K., HASEGAWA M.;
RT "Molecular cloning and characterization of the genes encoding the
RT immunoreactive major cell-surface proteins of Porphyromonas
RT gingivalis".
RL FEMS MICROBIOL. LETT. 120:23-30(1994).
DR EMBL; D28771; D1006511; -.
FT SIGNAL.
FT SIGNAL. 1 15 POTENTIAL.
FT CHAIN 50 563 CELL SURFACE PROTEIN.
SQ SEQUENCE 563 AA; 60786 MW; 0036195A CRC32;

Query Match 19.8%; Score 85; DB 2; Length 563;
Best Local Similarity 31.8%; Pred.No. 1.38e+00;
Matches 14; Conservative 11; Mismatches 18; Indels 1; Gaps 1;

Db 497 LGFNWNPVLPDPPSPENPNPNPDPEGTPVPTDPEQPLPD 540
| | | | | | | | | | | | | | | | | | | | |
Qy 12 LSFWDKINPAGGPDHGTGSLDPTND-PVPTLPAEPCSPFFQ 54

RESULT 10
ID P88150 PRELIMINARY; PRT; 498 AA.
AC P88150;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

```

Query Match 19.1%; Score 82; DB 14; Length 302;  
Best Local Similarity 42.9%; Pred. No. 3.44e+00;  
Matches 18; Conservative 9; Mismatches 9; Indels

Db 65 RAYPTRDPDHPGCGSLDHPGNPAQAGLPS-p-VPYAPL 104  
 : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Ov 18 KIWPAGGPB-D-HGTVPGSDDNDTPV-PT-LRAEPCSPFOL 55

RESULT 12  
ID O44776 PRELIMINARY; PRT; 408 AA.

DT 01-JUN-1998 (TREMBREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)  
DE F33D11.2 PROTEIN.

US CAENORHABDITIS ELEGANS.  
OC. EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;  
OC. RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PLODERINAE; CAENORHABDITIS.

[1]  
SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KRSTEN J., LAISTER N., LATRILLE P.,

RA LIGHTNING J., LLOYD C., McMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULISTO J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATFARSTON P.

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "9.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
RT *elegans*.";   
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;  
RA SAMMONS L., WOHLDMANN P., MULLEN G.;

RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF039720; G2773176; -.  
 RA WATERSTON R.;  
 RC STRAIN-BRISTOL N2;  
 RP SEQUENCE FROM N.A.  
 [3]  
 RN

```

SQ SEQUENCE 408 AA; 45741 MW; B022DEF4 CRC32;
Query Match 19.1%; Score 82; DB 5: Length
Best Local Similarity 26.7%; Pred. No. 3.44e+00;
Matches 12; Conservative 11; Mismatches 20; Indels 0
Db 40 KKFTYGAF-WFTIWEHGNGKSPDSTETATPLPPSPA-PVPT 82
: : : : : : : : : : : : : : : : : : : : : : : : :
QY 7 RLRAFLSFFWDKINQAPGEPDHGTGSLDNDPTVPTLAEPCSP 51
: : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	13	
ID	Q36155	
AC	Q36155	
DT	01-NOV-1996	(TREMBLEUR, 01, CREATED)
DT	01-NOV-1996	(TREMBLEUR, 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998	(TREMBLEUR, 08, LAST ANNOTATION UPDATE)

**THE UNIVERSITY OF CHICAGO**



\*\*\*\*\*

WQ5RELH (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:41:20 1999; MasPar time 9.94 Seconds  
115.582 Million cell updates/sec

Tabular output not generated.

Title: >US-09-099-053-2

Description: (56-109) from US09099053.pap (3 of 6)

Perfect Score: 399  
Sequence: 1 FLALYDFTARGBELSVRRG.....ARRLSGQPSAGLVPITHYAK 54

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 25.973; Variance 96.013; scale 0.271

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	399	100.0	488	39	W89248 Human SAD.	5.69e-33
2	114	28.6	298	15	R84183 Megakaryocyte kinase	1.06e-02
3	114	28.6	505	14	R85929 Protein tyrosine-kinase	1.06e-02
4	114	28.6	505	8	R41941 PTK gene LptK-2 prod.	1.06e-02
5	113	28.3	536	8	R39706 Human pp60 c-src prot	1.32e-02
6	111	27.8	60	22	W07876 Residues 81-140 of ch	2.04e-02
7	111	27.8	533	8	R39705 Chicken pp60 c-src pr	2.04e-02
8	109	27.3	834	22	W13506 Human signal mediator	3.16e-02
9	94	23.6	1146	3	R15157 Abelson Related Gene,	7.77e-01
10	94	23.6	1182	3	R15157 Abelson Related Gene,	7.77e-01
11	93	23.3	59	25	W34228 SH3 domain B of mouse	9.57e-01
12	93	23.3	182	25	W05402 Human clone 53 protein	9.57e-01
13	93	23.3	788	25	W05393 Mouse SH3P12 protein..	9.57e-01
14	91	22.8	126	39	W73534 Lymphoid cell protein	1.45e+00
15	90	22.6	48	9	R41126 Vector pHEZNeopt ju	1.79e+00
16	90	22.6	635	33	W48898 Candida albicans Caba	1.79e+00

17	87	21.8	318	36	W69429 Human secreted protei	3.31e+00
18	86	21.6	303	29	W42071 Human Crk-like protei	4.06e+00
19	84	21.1	303	17	R77439 Mouse CRKL protein..	6.10e+00
20	84	21.1	420	19	W06490 Beta-1-4-galactosyltr	6.10e+00
21	84	21.1	821	7	R35451 Mouse eps8.	6.10e+00
22	83	20.8	415	35	W71595 Murine tyrosine phosph	7.46e+00
23	82	20.6	620	17	R94535 ITK tyrosine kinase.	9.13e+00
24	81	20.3	620	17	R94535 ITK tyrosine kinase.	9.13e+00
25	81	20.3	217	21	W14004 Human GRB2.	1.12e+01
26	81	20.3	217	29	W42070 Growth factor recepto	1.12e+01
27	81	20.3	217	16	R85918 Human GRB-2.	1.12e+01
28	81	20.3	402	32	W59642 Amino acid sequence o	1.12e+01
29	81	20.3	970	15	R85089 EPH-like receptor pro	1.12e+01
30	81	20.3	994	24	W26366 Mouse Nuk tyrosine ki	1.12e+01
31	81	20.3	994	15	R87018 Receptor tyrosine kin	1.12e+01
32	80	20.1	176	13	R1943 Grb3-3 protein..	1.36e+01
33	80	20.1	217	15	R84636 Grb2 protein.	1.36e+01
34	80	20.1	317	5	R36061 Growth Factor Recepto	1.36e+01
35	80	20.1	403	32	W59641 Amino acid sequence o	1.36e+01
36	80	20.1	621	7	R37870 Soluble KEX2 protease	1.36e+01
37	80	20.1	622	4	R20469 Soluble yscf.	1.36e+01
38	80	20.1	712	4	R20468 KEX2 endopeptidase w1	1.36e+01
39	80	20.1	814	27	W25685 KEX2 endopeptidase.	1.36e+01
40	79	19.8	814	1	P90681 Eph-related PTK Cck5.	1.66e+01
41	79	19.8	995	13	R75712 Eph-related PTK Cck5+	1.66e+01
42	78	19.5	537	10	R46964 Peptide with Ikaros p	2.03e+01
43	77	19.3	63	20	W01472 Agammaglobulinaemia t	2.47e+01
44	77	19.3	442	20	W06709 Human haematopoietic-	2.47e+01
45	77	19.3	659	17	R94534 BTK tyrosine kinase.	2.47e+01

## ALIGNMENTS

RESULT 1

ID	W89248	standard; Protein; 488 AA.
AC	W89248	10-MAR-1999 (first entry)
DE	Human SAD.	
KW	PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;	
KW	type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;	
KW	neurodegenerative disease; neuronal survival; Alzheimer's disease;	
KW	Parkinson's disease; Huntington's disease.	
OS	Homo sapiens.	
PN	W09849317-A2.	
PD	05-NOV-1998.	
PF	27-APR-1998; U08439.	
PR	23-OCT-1997; US-063595.	
PR	28-APR-1997; US-044428.	
PR	20-MAY-1997; US-047222.	
PR	11-JUN-1997; US-049477.	
PR	11-JUN-1997; US-049756.	
PR	18-JUN-1997; US-049914.	
PA	(SUGEN) SUGEN INC.	
PI	App H, Clary D, Courtneidge SA, Hui TH, Jallal B,	
PI	Markby D, Onrust S, Peles E, Plozman GD;	
DR	WPI; 99-009434/01.	
DR	N-PSDB; V81743.	
PT	New nucleic acid encoding specific protein tyrosine phosphatases -	
PT	useful for identifying specific modulators for treatment and	
PT	prevention of cancer and neurodegenerative disease	
PS	Claim 2; Page 154-155; 193pp; English.	
CC	The present invention describes isolated, enriched or purified nucleic	
CC	acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The	
CC	present sequence represents human SAD. The above proteins, other than	
CC	ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify	
CC	substances that modulate their activity (i.e. agonists and antagonists,	
CC	including NBP) in vivo or in vitro. These substances are used to treat	
CC	or prevent diseases associated with abnormal signal transduction	
CC	pathways that involve the proteins, particularly cancer (e.g. leukaemia	
CC	and lymphoma), white modulators of ALK-7 (which is a type I receptor	
CC	serine/threonine kinase) are used to promote neuronal survival,	
CC	particularly for treating Alzheimer's, Parkinson's or Huntington's	
CC	diseases. Nucleic acid fragments of the polynucleotides encoding the	





CC library, and the bound enantiomer is then produced. The method can also  
 CC be used to produce L-oligonucleotides (such as T44064) and D-amino acid  
 CC peptide which bind to target L-MM's and for identifying an L-amino acid  
 CC peptide which binds a D-amino acid peptide of interest. The method is  
 CC used for identifying MM's of NNH (such as peptides, proteins,  
 CC oligonucleotides, sugars, carbohydrates, phospholipids, RNA-protein or  
 CC protein-lipid complexes) which are ligands for other target MM's, e.g.  
 CC DNA, hormones, and antibodies. The method is especially useful for  
 CC identifying D-peptides and L-nucleic acids (from naturally occurring  
 CC L-peptides and D-nucleic acids) for use as drugs. These compounds maybe  
 CC superior drugs to the corresponding compounds of NH since they are not  
 CC good substrates for naturally occurring proteases and nucleases and do  
 CC not elicit an efficient immune response.

SQ Sequence 60 AA;

Query Match            27.8%; Score 111; DB 23; Length 60;  
 Best Local Similarity 33.3%; Pred.No. 2.04e+02;  
 Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2

Db fvalydyesrtetdlfkfgerlqivntegdwlahslttgqt-gypisnyva 58  
 :|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
 QY 56 FLALYDFTARCGGELSVRGRDLCALEEG-GGYIFARRLSGGPSAGLVPITHVA 108  
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

RESULT 7  
 ID R39705 standard; Protein: 533 AA.  
 AC R39705;  
 DE 23-DEC-1993 (first entry)  
 DT Chicken pp60 c-src protein.  
 KW Endothelial; tyrosine kinase protein; pp60 c-src; ss.  
 OS Gallus gallus.  
 PN WO9314193-A.  
 PP 05-JAN-1993; US000445.  
 PR 06-JAN-1992; US-820011.  
 PA {OYYA} UNIV YALE.  
 PI Bell L, Lutheringer DJ, Madri JA, Warren SL;  
 DR WPI; 93-243209/30.  
 P-PSDB; R39705.  
 PT Genetically engineered endothelial cells - which exhibit enhanced  
 PT cell migration, urokinase-type plasminogen activator activity,  
 PT and reduced mononuclear cell adhesion and fibronectin prodn  
 PS Disclosure; Page 64-66; 91pp; English.  
 CC The DNA encoding a portion or (more preferably) the entire pp60  
 CC c-src polypeptide (Given in Q45687) is used to transform endothelial  
 CC cells, transformed cells produce increased amounts of pp60 c-src and  
 CC have improved therapeutic properties. They migrate at faster rates  
 CC than non-transformed counterparts; have an enhanced ability to  
 CC inhibit the formation of thrombi and/or dissolve thrombi once they  
 CC have formed and exhibit reduced mononuclear cell adhesion. They can  
 CC also be used to improve the success of surgical procedures such as  
 CC coronary angioplasty, heart bypass surgery, vessel graft and stent  
 CC implantation..  
 SQ Sequence 533 AA;

Query Match            27.8%; Score 111; DB 8; Length 533;  
 Best Local Similarity 33.3%; Pred.No. 2.04e+02;  
 Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2

Db fvalydyesrtetdlfkfgerlqivntegdwlahslttgqt-gypisnyva 138  
 :|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
 QY 56 FLALYDFTARCGGELSVRGRDLCALEEG-GGYIFARRLSGGPSAGLVPITHVA 108  
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

RESULT 8  
 ID WL1506 standard; Protein: 834 AA.  
 AC WL1506;  
 DT 22-JUN-1997 (first entry)  
 DE Human signal mediator protein encoded by HEF1 CDNA.  
 KW Signal mediator protein; SMP; human enhancer of filamentation;  
 KW HEF1; cell morphology; neoplasia; SH3; SH2; pseudohyphal budding;  
 yeast.  
 OS Homo sapiens.

DT	12-FEB-1992 (first entry)
DE	Abelson Related Gene, B transcript.
KW	Arg; diagnosis; therapy; tumour; abl proto-oncogene.
OS	Homo sapiens.
PN	U57559029-A.
PD	22-OCT-1991.
PF	22-OCT-1991; 559029.
PR	30-JUL-1990; US-559029.
PT	(USSH ) NAT INST OF HEALTH.
PI	Kruh G, Aronson SA, King CR;
DR	WPI: 91-353425/48.
DR	N-PSDB; Q14937.
PT	Novel human gene related to abl proto-oncogene - des-
PT	cribed as "Abelson Related Gene", arg, useful for tumour diag-
PT	therapy
PS	Disclosure; Fig 5D; English.
CC	The human gene encoding this protein is closely rela-
CC	ted to the human gene encoding the protein product of
CC	distinct from the abl proto-oncogene and is a member
CC	of the family of genes. Arg is expressed as a
CC	kinase encoding family of genes. Arg is expressed as a
CC	By analogy with c-abl, the alternative 5' arg sequen-
CC	designated A (Q14936) and B and it is assumed that t
CC	to the arg second exon.
CC	The amino acid sequence is represented as found in t
CC	Sequence 1182 AA;
SQ	

```
Query Match      23.6%; Score 94; DB 3; Length 1182;
Best Local Similarity 46.2%; Pred. No. 7.77e-01;
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
```

Db 112 fvalydfvasgdntlsitkgeklrvl 137  
|:||||| | ||: |:| |  
QY 56 FLALYDFTARCGGELSVRRGDRLCAL 81

```

RESULT 11
ID W34228 standard; Protein; 59 AA.
AC W34228;
DT 20-FEB-1998 (first entry)
DE SH3 domain B of mouse SH3p12 protein.
KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process.
OS Mus musculus.
PN WO9631625-A1.
PD 10-OCT-1996.
PF 04-APR-1996; U04454.
PR 03-APR-1996; US-630915.
PR 07-APR-1995; US-417872.
PA (CYTO-) CYTOGEN CORP.
PA (UNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
PT Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
CLAIM 55; Fig 10a; 174pp; English.
CC W34220-W34253 represent Src-homology region 3 (SH3) domains from human
CC and mouse proteins, that can be used in the method of the invention. This
CC sequence represents a SH3 domain of the mouse SH3p12 protein (see WO5283

```

Query Match 23.6%; Score 94; DB 3; Length 1146;  
Best Local Similarity 46.2%; Pred. No. 7.77e-01;  
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

```

76 fvalydfvasgdntlsitkgeklrvl 101
   |:||||| | ||: |:| |
56 FLALYDFTARC GGELSVRRGDLCAL 81

```

RESULT 10  
0 RI5157 standard; Protein; 1182 AA.  
1 RI5157;



Job time : 128 secs.

```
PD 15-DEC-1998.
PF 28-MAY-1997; 138905.
PR 28-MAY-1997; JP-138905.
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (MITU ) MITSUBISHI CHEM CORP.
DR WPI; 99-099029/09.
DR N-PSDB; V62889.
PT New immunosuppressant DNA and protein - useful for inhibition and
PT treatment of autoimmune diseases caused by lymphoid cell
PT protein-tyrosine kinase analogues
PS Claim 1; Page 4-5; 6pp; Japanese.
CC This sequence is the Lymphoid cell protein-tyrosine kinase (Lck) SH3
CC domain of the invention. The DNA and protein are useful as immune
CC suppressants, and are useful for inhibition and treatment of autoimmune
CC diseases caused by Lck analogues.
SQ Sequence 126 AA;

Query Match 22.8%; Score 91; DB 39; Length 126;
Best Local Similarity 30.2%; Pred. NO. 1.45e+00;
Matches 16; Conservative 15; Mismatches 21; Indels 1; Gaps 1;

Db 67 lalhsyepshdgdldgfekgqrlileqsgwkwkagslttgg-egfipfxfvak 118
:|: : : |: : : |: : |: : |: : |: : |: : |: : |: : |: : |:
QY 57 LALYDFTARCGGELSVRRGDRLCALCEGGGYIFARRLSQGPSAGLVPIYTHVAK 109

RESULT 15
ID R47126 standard; Protein; 48 AA.
AC R47126;
DE 11-JUL-1994 (first entry)
DT Vector pHF2prNeoDr junction fragment.
KW MVM pause signal sequence; cassette; neo gene; MCl promoter;
KW diphtheria toxin A; DT-A; bluescript DNA; PCK promoter;
KW homologous recombination; embryonic stem cell.
OS Synthetic.
FH Key Location/Qualifiers
FT peptide 1..27
/note= "fyn Aderived amino acids"
FT peptide 28..32
/note= "Linker amino acids"
FT peptide 33..48
/note= "Trp-lacZ amino acids"
PN J05328966-A.
PD 14-DEC-1993.
PF 26-MAY-1992; 158592.
PR 26-MAY-1992; JP-158592.
PA (RIKA ) RIKAGAKU KENKYUSHO.
DR WPI; 94-021918/03.
PT Selection of cells containing diphtheria toxin A fragment - using
PT a cassette containing neo gene, PCK promoter, MVM pause signal
PT and Bluescript DNA
PS Disclosure; Fig 4; 7pp; Japanese.
CC This sequence is encoded by a fragment of the vector pHF2prNeoDr. This
CC vector contained a cassette, containing the neo gene, MVM pause signal
CC sequence, MCl promoter, a diphtheria toxin A (DT-A) fragment and
CC bluescript DNA. The neo gene has a PCK promoter just upstream of it,
CC and the MVM pause signal and bluescript DNA are inserted between the
CC neo gene and the MCl promoter. This construct was used in the
CC selection of a cell containing a genomic DNA gene which is deleted
CC or reduced by homologous recombination with DT-A. The cassette is
CC preferably expressed in embryonic stem cells.
SQ Sequence 48 AA;

Query Match 22.6%; Score 90; DB 9; Length 48;
Best Local Similarity 39.3%; Pred. NO. 1.79e+00;
Matches 11; Conservative 9; Mismatches 7; Indels 1; Gaps 1;

Db 4 fvalydyeaertddlsfekgkxf-qmed 30
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 56 FLALYDFTARCGGELSVRRGDRLCALCE 83

Search completed: Thu May 20 12:43:28 1999
```



```

targeted disruption.
#accession A56040
##status preliminary
##molecule_type mRNA
##residues 1-496 #label KOH
##cross-references GB:D26186; NID:g529072; PID:d1005873; PID:g529073
GENETICS
#map_position 2
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
kinases; protein kinase homology; SH3 homology
ATP; phosphotransferase
FEATURE
62-111 #domain SH3 homology #label SH3\
232-491 #domain protein kinase homology #label KIN\
240-248 #region protein kinase ATP-binding motif
SUMMARY
#length 496 #molecular-weight 55593 #checksum 301
Query Match 59.7%; Score 278; DB 2; Length 496;
Best Local Similarity 72.2%; Pred. No. 4.07e-41;
Matches 39; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
Db 60 FLALYDFTARCABELSVGGDRLYALKEGDYIFAOIRLGGPPSTGLVPVTVLAK 113
QY 56 FLALYDFTARCAGGELSVRRGDRLCALKEEGGYIFARRLSQPSAGLVPIITHVA 109

RESULT 3 TVHAST #type complete
ENTRY protein-tyrosine kinase (EC 2.7.1.112) stk - Hydra attenuata
TITLE #formal_name Hydra attenuata
ORGANISM 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
DATE 05-Sep-1997
ACCESSIONS A34094
REFERENCE Bosch, T.C.G.; Unger, T.F.; Fisher, D.A.; Steele, R.E.
Mol. Cell. Biol. (1989) 9:4141-4151
#authors
#journal
#title Structure and expression of STK, a src-related gene in the
simple metazoan Hydra attenuata.
#cross-references MUID:90066418
#accession A34094
##molecule_type mRNA
##residues 1-509 #label BOS
##cross-references GB:M25245; NID:g159273; PID:g159274
GENETICS
#gene stk
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3-homology
ATP; autophosphorylation; phosphoprotein; phosphotransferase;
transforming protein; tyrosine-specific protein kinase
FEATURE
66-115 #domain SH3 homology #label SH3\
126-218 #domain SH2 homology #label SH2\
238-497 #domain protein kinase homology #label KIN\
246-254 #region protein kinase ATP-binding motif\
268 #active_site Lys #status predicted\
390 #binding_site phosphate (Tyr) (covalent) (by
autophosphorylation) #status predicted
SUMMARY
#length 509 #molecular-weight 56885 #checksum 8721
Query Match 33.6%; Score 134; DB 1; Length 509;
Best Local Similarity 40.7%; Pred. No. 2.25e-10;
Matches 22; Conservative 11; Mismatches 19; Indels 2; Gaps 2;
Db 64 FVALYDYEARISELSFKKGERLQINTADGDWYARSLTN-SEGVIPTTYVA 116
QY 56 FLALYDFTARCAGGELSVRRGDRLCALKEEGGYIFARRLSQPSAGLVPIITHVA 108

RESULT 4
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) src 2 - African clawed
frog
ALTERNATE_NAMES kinase-related transforming protein (src); kinase-related
transforming protein (src)

transforming protein (src) 2
#formal_name Xenopus laevis #common_name African clawed frog
30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change
20-Mar-1998
ACCESSIONS B34104; I51563
REFERENCE A34104
#authors Steele, R.E.; Unger, T.F.; Mardis, M.J.; Fero, J.B.
#journal J. Biol. Chem. (1989) 264:10649-10653
#title The two Xenopus laevis SRC genes are co-expressed and each
produces functional pp(60src).
#cross-references MUID:89278134
#accession B34104
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-532 #label STE
##cross-references GB:M23422; GB:J04822; NID:g214796; PID:g214797
REFERENCE I51563
#authors Steele, R.E.
#journal Nucleic Acids Res. (1985) 13:1747-1761
#title Two divergent cellular src genes are expressed in Xenopus
laevis.
#cross-references MUID:85215578
#accession I51563
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 439-492 #label ST2
##cross-references GB:M30858; NID:g214799; PID:g555569
GENETICS
#gene src
#introns 464/1
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
KEYWORDS ATP; autophosphorylation; phosphoprotein; phosphotransferase;
tyrosine-specific protein kinase
FEATURE
87-136 #domain SH3 homology #label SH3\
147-244 #domain SH2 homology #label SH2\
264-522 #domain protein kinase homology #label KIN\
272-280 #region protein kinase ATP-binding motif
SUMMARY
#length 532 #molecular-weight 59736 #checksum 7595
Query Match 31.8%; Score 127; DB 2; Length 532;
Best Local Similarity 38.9%; Pred. No. 4.93e-09;
Matches 21; Conservative 13; Mismatches 18; Indels 2; Gaps 2;
Db 85 FVALYDYESTETDLSFRKGERLQIVNTEGDWLRSLSGQT-GYIPSNVA 137
QY 56 FLALYDFTARCAGGELSVRRGDRLCALKEEG-GGYIFARRLSQPSAGLVPIITHVA 108

RESULT 5
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) src 1 - African clawed
frog
ALTERNATE_NAMES kinase-related transforming protein (src); kinase-related
transforming protein (src)
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change
20-Mar-1998
ACCESSIONS A34104; I51564
REFERENCE A34104
#authors Steele, R.E.; Unger, T.F.; Mardis, M.J.; Fero, J.B.
#journal J. Biol. Chem. (1989) 264:10649-10653
#title The two Xenopus laevis SRC genes are co-expressed and each
produces functional pp(60src).
#cross-references MUID:89278134
#accession A34104
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-532 #label STE
##cross-references GB:M24704; GB:J04822; NID:g214804; PID:g214805
REFERENCE I51564
#authors Steele, R.E.; Chosn, R.; Ral, B.B.A.; Winokur, S.T.; Unger,

```

241-499	#domain protein kinase homology #label KIN\
249-257	#region protein kinase ATP-binding motif\
2	#modified site myristylated amino end (Gly) (in mature form) #status predicted
SUMMARY	#length 507 #molecular-weight 58011 #checksum 4536
Query Match	29.8%; Score 119; DB 2; Length 507;
Best Local Similarity	35.8%; Pred. No. 1.56e-07;
Matches	19; Conservative 15; Mismatches 18; Indels 1; Gaps 1;
Ddb	65 VALYDEPHDGLKQEKLVLESEGEWRAQSLLTQO-EGLTPHNFMVAM 116 :    :  : ::  :: :     :  :: :  :: :  :: :
OQ	57 LALYDFTRCGGELSYYRRGDRLCALCEGGGYIFARRLSGQPSAGLVPITHVAK 109 :    :  : ::  :: :     :  :: :  :: :  :: :
RESULT	7
ENTRY	I51592 #type complete
TITLE	p59(xfyn) - Xiphophorus helleri
ORGANISM	#formal_name Xiphiphorus helleri
DATE	04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 10-Jul-1998
ACCESSIONS	I51592
REFERENCE	I51592 preliminary; translated from GB/EMBL/DDBJ
#authors	Hannig, G.; Ottillie, S.; Schartl, M.
#Journal	Oncogene (1991) 6:361-369
#title	Conservation of structure and expression of the c-yes and fyn genes in lower vertebrates.
#cross-references	MUID:91187435
#accession	I51592
#status	preliminary; translated from GB/EMBL/DDBJ
GENETICS	#molecule_type mRNA
#residues	1-537 #label HAN
#cross-references	EMBL:X54971; NID:g64481; PID:g64482
#gene	xfin
CLASSIFICATION	#superfamily protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH3 homology
FEATURE	#domain SH3 homology #label SH3\
89-138	#domain protein kinase homology #label KIN
269-527	#length 537 #molecular-weight 60447 #checksum 621
SUMMARY	
Query Match	29.6%; Score 118; DB 2; Length 537;
Best Local Similarity	38.9%; Pred. No. 2.39e-07;
Matches	21; Conservative 11; Mismatches 20; Indels 2; Gaps 2;
Ddb	87 FVALDYERTEDDLSFRGERFQILNSTEGDMWDARSITT-GGSYIIPSNYA 139 :    :  : ::  :: :  :: :  :: :  :: :  :: :
OQ	56 FLALYDFTRCGGELSYYRRGDRLCALCEGGGYIF-ARRLSGQPSAGLVPITHVA 108 :    :  : ::  :: :  :: :  :: :  :: :  :: :
RESULT	8
ENTRY	S26420 #type complete
TITLE	protein-tyrosine kinase (EC 2.7.1.112) src - Rous sarcoma virus
ORGANISM	#formal_name Rous sarcoma virus
DATE	06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 19-Dec-1997
ACCESSIONS	S26420; S20676
REFERENCE	S26417
#authors	Kashuba, V.I.; Rynditch, A.V.; Dostalova, V.; Hlozaneck, I.; Zubak, S.V.; Kavsan, V.M.
#submission	submitted to the EMBL Data Library, September 1992
#description	Molecular cloning and DNA sequence analysis of duck-adapted variant of Rous sarcoma virus (da Pr-RSV-C).
#accession	S26420
#status	preliminary
#molecule_type	DNA
#residues	1-526 #label KAS
#cross-references	EMBL:X68524; NID:g61903; PID:g61907
REFERENCE	S20676
#authors	Kashuba, V.I.; Serge, Z.V.; Rynditch, A.V.; Kavsan, V.M.; Hlozaneck, I.



```

#submission      submitted to the EMBL Data Library, March 1990
#accession      S20676
#status          preliminary
#title           protein-tyrosine kinase src; protein kinase
#molecule_type DNA
#residues        1-526 ##label KA2
#cross-references EMBL:X51861; NID:g61896; PID:g61897

GENETICS
#gene            src
#classification  #superfamily protein-tyrosine kinase src; protein kinase
#keywords         homology; SH2 homology; SH3 homology
#feature          ATP; phosphotransferase
#feature         88-137   #domain SH3 homology #label SH3\
#feature         148-245   #domain SH2 homology #label SH2\
#feature         265-523   #domain protein kinase homology #label KIN\
#feature         273-281   #region protein kinase ATP-binding motif
#summary          #length 526 #molecular-weight 59212 #checksum 7994

Query Match      28.8%; Score 115; DB 2; Length 526;
Best Local Similarity 33.3%; Pred. No. 8.52e-07;
Matches          18; Conservative 16; Mismatches 18; Indels 2; Gaps 2;

Db      86 FVALDYESTWTDLSFKKERLQVNNTEGYWMLAHSLTGQT-GYIPSNYYA 138
1:||||: :: ||| :|:|:| ::| ||| :|:| :| :| ||
QY      56 FLALYDFTRCGGELSVRRGDRLCALEE-EGGGYI-FARRLSGGPSAGLVPIITHVA 108

RESULT          9
ENTRY           I38396      #type complete
TITLE           protein-tyrosine kinase (EC 2.7.1.112) FRK - human
ALTERNATE_NAMES FYN-related kinase (FRK)
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change
                09-Apr-1998
ACCESSIONS      I38396
REFERENCE        Lee, J.; Wang, Z.; Luoh, S.M.; Wood, W.I.; Scadden, D.T.
#authors         Gene (1994) 138:247-251
#journal          Cloning of FRK, a novel intracellular SRC-like tyrosine
#title            kinase-encoding gene.
#cross-references EMBL:U00803; NID:g392887; PID:g392888
#accession       I38396
#status          preliminary
#molecule_type mRNA
#residues        1-505 ##label RES
#cross-references EMBL:U00803; NID:g392887; PID:g392888

GENETICS
#gene            GDB:FRK
#cross-references GDB:355675
#map_position     4q35-4q35
#classification  #superfamily protein-tyrosine kinase src; protein kinase
#keywords         homology; SH2 homology; SH3 homology
#feature          ATP; phosphotransferase
#feature         49-105   #domain SH3 homology #label SH3\
#feature         232-494   #domain protein kinase homology #label KIN\
#feature         240-248   #region protein kinase ATP-binding motif
#summary          #length 505 #molecular-weight 59254 #checksum 9379

Query Match      28.6%; Score 114; DB 2; Length 505;
Best Local Similarity 46.2%; Pred. No. 1.30e-06;
Matches          8; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db      47 FVALFDYQARTADLSFRAGDKLOVLDTLHEGWFFARHL 85
1:|:|:|: | | :|:| | | | | | :| :| :| :| :| :|
QY      56 FLALYDFTRCGGELSVRRGDRLCALEE-GGGYIFARRL 93

RESULT          10
ENTRY           I49552      #type complete
TITLE           protein-tyrosine kinase (EC 2.7.1.112) bsk/tyk - mouse
ALTERNATE_NAMES intestinal tyrosine kinase
ORGANISM        #formal_name Mus musculus #common_name house mouse

```

```

271-529      #domain protein kinase homology #label KIN\
279-287      #region protein kinase ATP-binding motif
SUMMARY      #length 539 #molecular-weight 60720 #checksum 8770

Query Match      28.18; Score 112; DB 2; Length 539;
Best Local Similarity 33.38; Pred. No. 3.00e-06;
Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

Db 89 FTALYNDARTEDDLFRKGEKHIINSSEGDWWEARSLT-GSTGYIPSNYYA 141
   |||||::||::||::||::||::||::||::||::||::||::||::||::||
QY 56 FLALYDFTARCGELSVRRGRLCALEEGGYIF-ARRLSGQPSAGLVPITHVA 108
   |||||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 12
ENTRY      S24547      #type complete
TITLE      protein-tyrosine kinase (EC 2.7.1.112) fgr - rat
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change
          19-Dec-1997
ACCESSIONS S24547; PT0200
REFERENCE   Yue, C.C.
#authors   Yue, C.C.
#submission submitted to the EMBL Data Library, December 1990
#accession S24547
#status    preliminary
#molecule_type mRNA
#residues  1-517 #label YUE
#cross-references EMBL:X57018; NID:G56145; PID:G56146
REFERENCE   PT0196
#authors   Yue, C.C.
#journal   Mol. Immunol. (1991) 28:399-408
#title     Novel putative protein kinase clones from a rat large
           granular lymphocyte tumor cell line.
#cross-references MUID:91287726
#accession PT0200
#molecule_type mRNA
#residues  371-427 #label YU2
#experimental_source lymphocyte cell line

GENETICS
#gene      FGR
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
               homology; SH2 homology; SH3 homology
KEYWORDS     ATP; phosphotransferase; tyrosine-specific protein kinase
FEATURE
72-121      #domain SH3 homology #label SH3\
132-229      #domain SH2 homology #label SH2\
245-507      #domain protein kinase homology #label KIN\
257-265      #region protein kinase ATP-binding motif
SUMMARY      #length 517 #molecular-weight 58792 #checksum 9498

Query Match      27.88; Score 111; ..DB.2.. Length 517; ..
Best Local Similarity 38.28; Pred. No. 4.54e-06;
Matches 21; Conservative 12; Mismatches 18; Indels 4; Gaps 3;

Db 70 FVALDYEARTGDDLTFTTKEKHIINLNTF-YDWWEARSLSS-GRGTGYVPSNYYA 122
   |||||::||::||::||::||::||::||::||::||::||::||::||::||
QY 56 FLALYDFTARCGELSVRRGRLCALEEGGYIF-ARRLSGQPSAGLVPITHVA 108
   |||||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 13
ENTRY      TVCHS      #type complete
TITLE      protein-tyrosine kinase (EC 2.7.1.112) src - chicken
ALTERNATE_NAMES kinase-related transforming protein src
ORGANISM     #formal_name Gallus gallus #common_name chicken
DATE       19-Feb-1984 #sequence_revision 07-Oct-1994 #text_change
          07-Nov-1997
ACCESSIONS A00630; A41256; A32432; C35650
REFERENCE   Takeya, T.; Hanafusa, H.
#authors   Takeya, T.; Hanafusa, H.
#journal   Cell (1983) 32:881-890
#title     Structure and sequence of the cellular gene homologous to the
           RSV sec gene and the mechanism for generating the
           transforming virus.

```

```

#cross-references MUID:83155664
#accession A00630
#molecule_type DNA
#residues  1-500, 'R', 502-533 #label TAK
#note      the authors translated the codons AAC and CAG for
           residues 301 and 526 as Thr and Glu, respectively
REFERENCE   A41256
#authors   Dorai, T.; Levy, J.B.; Kang, L.; Brugge, J.S.; Wang, L.H.
#journal   Mol. Cell. Biol. (1991) 11:4165-4176
#title     Analysis of cDNAs of the proto-oncogene c-src: heterogeneity
           in 5' exons and possible mechanism for the genesis of the
           3' end of v-src.
#accession A41256
#molecule_type mRNA
#residues  484-533 #label DOR
#cross-references GB:S43579
#note      the authors translated the codon CAG for residue 527 as
           Glu
REFERENCE   A32432
#authors   Shenoy, S.; Choi, J.K.; Bagrodia, S.; Copeland, T.D.; Maller,
           J.L.; Shalloway, D.
#journal   Cell (1989) 57:763-774
#title     Purified maturation promoting factor phosphorylates pp60
           (c-src) at the sites phosphorylated during fibroblast
           mitosis
#cross-references MUID:89249341
#accession A32432
#molecule_type protein
#residues  2-88 #label SHE
#note      34-Thr, 46-Thr, and 72-Ser are phosphorylated during
           mitosis
REFERENCE   A35650
#authors   Dorai, T.; Wang, L.H.
#journal   Mol. Cell. Biol. (1990) 10:4068-4079
#title     An alternative non-tyrosine protein kinase product of the
           c-src gene in chicken skeletal muscle.
#cross-references MUID:90318371
#accession C35650
#molecule_type mRNA
#residues  1-182, 'DP', 185, 'IPLPSCIC', #label DO2
#cross-references GB:M57290; NID:G212703; PID:G212706
#note      alternatively spliced mRNA exclusively replaces the long
           form in skeletal muscle shortly before hatching
           this ORF appears not to be translated
#note      A90838
REFERENCE   Takeya, T.; Hanafusa, H.
#authors   Takeya, T.; Hanafusa, H.
#journal   Cell (1983) 34:319
#contents  annotation; erratum, correct translation of residue 526
GENETICS
#gene      src
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
               homology; SH2 homology; SH3 homology
KEYWORDS     alternative splicing; ATP; autophosphorylation; blocked amino
           end; lipoprotein; myristylation; phosphoprotein;
           phosphotransferase; proto-oncogene; tyrosine-specific
           protein kinase
FEATURE
88-137      #domain SH3 homology #label SH3\
148-245      #domain SH2 homology #label SH2\
265-523      #domain protein kinase homology #label KIN\
273-281      #region protein kinase ATP-binding motif\
2           #modified site myristylated amino end (Gly) (in mature
           form) #status predicted\
12, 48      #binding site phosphate (Ser) (covalent) (by protein
           kinase C) #status predicted\
17          #binding site phosphate (Ser) (covalent) (by protein
           kinase A) #status predicted\
34, 46      #binding site phosphate (Thr) (covalent) #status
           experimental\
72          #binding site phosphate (Ser) (covalent) #status
           structural\
295         #active site Lys #status predicted\
527         #binding site phosphate (Tyr) (covalent) #status

```

```
SUMMARY      #length 533 #molecular-weight 60023 #checksum 238
              predicted
Query Match   27.8%; Score 111; DB 1; Length 533;
Best Local Similarity 33.3%; Pred. No. 4.54e-06;
Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

Db 86 FVALYDYESRTETDLSFKKGERLQIVNTEGDWLAHSLTTGOT-GYIPSNVYA 138
|:||||:| :|| :||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
56 FLALYDFTARGGELSRRGRDLCALEEG-GGYIFARRLSGQPSAGLVPITHVA 108

RESULT 14
ENTRY   TVFVS2      #type complete
TITLE   protein-tyrosine kinase (EC 2.7.1.112) src - avian sarcoma
        virus S2
ORGANISM #formal_name avian sarcoma virus S2
DATE     31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change
        07-Nov-1997
ACCESSIONS B25375
REFERENCE  A25375
#authors   Ikawa, S.; Hagino-Yamagishi, K.; Kawai, S.; Yamamoto, T.;
           Toyoshima, K.
#journal   Mol. Cell. Biol. (1986) 6:2420-2428
#title     Activation of the cellular src gene by transducing
           retrovirus.
#cross-references MUID:87064539
#accession B25375
#molecule_type DNA
#residues  1-557 ##label IKA

GENETICS
#gene      src
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
               homology; SH2 homology; SH3 homology
KEYWORDS    ATP; autophosphorylation; blocked amino end; lipoprotein;
               myristylation; oncogene; phosphoprotein;
               phosphotransferase; transforming protein; tyrosine-specific
               protein kinase

FEATURE
88-137      #domain SH3 homology #label SH3\
148-245     #domain SH2 homology #label SH2\
265-523     #domain protein kinase homology #label KIN\
273-281     #region protein kinase ATP-binding motif\
2           #modified_site myristylated amino end (Gly) (in mature
           form) #status predicted\
295         #active_site Lys #status predicted\
416         #binding_site phosphate (Tyr) (covalent) (by
           autophosphorylation) #status predicted

SUMMARY      #length 557 #molecular-weight 62582 #checksum 7810
              predicted
Query Match   27.8%; Score 111; DB 1; Length 557;
Best Local Similarity 33.3%; Pred. No. 4.54e-06;
Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

Db 86 FVALYDYESRTETDLSFKKGERLQIVNTEGDWLAHSLTTGOT-GYIPSNVYA 138
|:||||:| :|| :||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
56 FLALYDFTARGGELSRRGRDLCALEEG-GGYIFARRLSGQPSAGLVPITHVA 108

RESULT 15
ENTRY   TVFVS1      #type complete
TITLE   protein-tyrosine kinase (EC 2.7.1.112) src - avian sarcoma
        virus S1
ORGANISM #formal_name avian sarcoma virus S1
DATE     31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change
        07-Nov-1997
ACCESSIONS A25375
REFERENCE  A25375
#authors   Ikawa, S.; Hagino-Yamagishi, K.; Kawai, S.; Yamamoto, T.;
           Toyoshima, K.
#journal   Mol. Cell. Biol. (1986) 6:2420-2428
#title     Activation of the cellular src gene by transducing
           retrovirus.
```

```
#cross-references MUID:87064539
#accession A25375
#molecule_type DNA
##residues 1-568 ##label IKA

GENETICS
#gene      src
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
               homology; SH2 homology; SH3 homology
KEYWORDS    ATP; autophosphorylation; blocked amino end; lipoprotein;
               myristylation; phosphoprotein; phosphotransferase;
               transforming protein; tyrosine-specific protein kinase

FEATURE
88-137      #domain SH3 homology #label SH3\
148-245     #domain SH2 homology #label SH2\
265-523     #domain protein kinase homology #label KIN\
273-281     #region protein kinase ATP-binding motif\
2           #modified_site myristylated amino end (Gly) (in mature
           form) #status predicted\
295         #active_site Lys #status predicted\
416         #binding_site phosphate (Tyr) (covalent) (by
           autophosphorylation) #status predicted

SUMMARY      #length 568 #molecular-weight 63632 #checksum 4430
              predicted
Query Match   27.8%; Score 111; DB 1; Length 568;
Best Local Similarity 33.3%; Pred. No. 4.54e-06;
Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

Db 86 FVALYDYESRTETDLSFKKGERLQIVNTEGDWLAHSLTTGOT-GYIPSNVYA 138
|:||||:| :|| :||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
56 FLALYDFTARGGELSRRGRDLCALEEG-GGYIFARRLSGQPSAGLVPITHVA 108

Search completed: Thu May 20 12:45:34 1999
Job time : 107 secs.
```

\*\*\*\*\*

W P E R F H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:45:53 1999; MasPar time 4.06 Seconds

Tabular output not generated. .... 375.926 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (56-109) from US09099053.pap (3 of 6)  
Perfect Score: 399  
Sequence: 1 FLALYDFATRCGSGELSVRRG.....ARRLSGQPSAGLVPITHVAK 54

Scoring table: PAM 150  
Gap 11

Searched: 77977 segs, 28258293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 36.365; Variance 52.725; scale 0.690

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	292	73.2	496	1 SRM_MOUSE	TYROSINE-PROTEIN KINAS	1.15e-51
2	134	33.6	509	1 STK_HYDAT	TYROSINE-PROTEIN KINAS	2.07e-12
3	127	31.8	531	1 SRC2_XENLA	TYROSINE-PROTEIN KINAS	7.23e-11
4	122	30.6	531	1 SRC1_XENLA	TYROSINE-PROTEIN KINAS	8.78e-10
5	119	29.8	507	1 LCK_CHICK	PROTO-ONCOGENE TYROSIN	6.29e-09
6	118	29.6	536	1 FYN_XIPHE	TYROSINE-PROTEIN KINAS	4.39e-08
7	114	28.6	505	1 SRC_HUMAN	TYROSINE-PROTEIN KINAS	7.10e-08
8	113	28.3	535	1 SRC_AVISR	TYROSINE-PROTEIN KINAS	1.85e-07
9	111	27.8	526	1 SRC_CHICK	PROTO-ONCOGENE TYROSIN	1.85e-07
10	111	27.8	532	1 SRC_AVIS	TYROSINE-PROTEIN KINAS	1.85e-07
11	111	27.8	557	1 SRC_AVIS	TYROSINE-PROTEIN KINAS	1.85e-07
12	111	27.8	568	1 SRC_AVIS	TYROSINE-PROTEIN KINAS	1.85e-07
13	111	27.8	587	1 SRC_AVIS2	TYROSINE-PROTEIN KINAS	2.98e-07
14	110	27.6	535	1 YRK_CHICK	PROTO-ONCOGENE TYROSIN	2.98e-07
15	110	27.6	544	1 YES_XIPHE	TYROSINE-PROTEIN KINAS	4.78e-07
16	109	27.3	526	1 HCK_HUMAN	TYROSINE-PROTEIN KINAS	4.78e-07
17	109	27.3	537	1 YES_XENLA	PROTO-ONCOGENE TYROSIN	7.66e-07
18	108	27.1	503	1 HCK_RAT	TYROSINE-PROTEIN KINAS	7.66e-07
19	108	27.1	524	1 HCK_MOUSE	TYROSINE-PROTEIN KINAS	7.66e-07
20	107	26.8	517	1 FGR_MOUSE	PROTO-ONCOGENE TYROSIN	1.23e-06
21	107	26.8	536	1 FYN_XENLA	PROTO-ONCOGENE TYROSIN	1.23e-06
22	106	26.6	533	1 FYN_MOUSE	PROTO-ONCOGENE TYROSIN	1.96e-06
23	106	26.6	536	1 FYN_HUMAN	PROTO-ONCOGENE TYROSIN	1.96e-06

24	105	26.3	541	1 YES_CHICK	PROTO-ONCOGENE TYROSIN	3.12e-06
25	104	26.1	526	1 SRC_RSVH1	TYROSINE-PROTEIN KINAS	4.96e-06
26	104	26.1	526	1 SRC_RSVVP	TYROSINE-PROTEIN KINAS	4.96e-06
27	103	25.8	540	1 SRCN_MOUSE	NEURONAL PROTO-ONCOGEN	7.87e-06
28	102	25.6	511	1 LYN_MOUSE	TYROSINE-PROTEIN KINAS	1.25e-05
29	102	25.6	511	1 LYN_RAT	TYROSINE-PROTEIN KINAS	1.25e-05
30	102	25.6	529	1 FGR_HUMAN	PROTO-ONCOGENE TYROSIN	1.25e-05
31	102	25.6	543	1 YES_HUMAN	PROTO-ONCOGENE TYROSIN	1.25e-05
32	100	25.1	533	1 FYN_CHICK	PROTO-ONCOGENE TYROSIN	3.10e-05
33	100	25.1	541	1 YES_MOUSE	PROTO-ONCOGENE TYROSIN	3.10e-05
34	98	24.6	504	1 BLK_HUMAN	TYROSINE-PROTEIN KINAS	7.65e-05
35	97	24.3	505	1 SRK1_SPOLA	TYROSINE-PROTEIN KINAS	1.20e-04
36	97	24.3	506	1 SRK4_SPOLA	TYROSINE-PROTEIN KINAS	1.20e-04
37	97	24.3	526	1 SRC_RSVSR	TYROSINE-PROTEIN KINAS	1.20e-04
38	96	24.1	528	1 YES_AVISY	TYROSINE-PROTEIN KINAS	1.87e-04
39	94	23.6	439	1 ABL_FSVHY	TYROSINE-PROTEIN KINAS	4.53e-04
40	94	23.6	1123	1 ABL_MOUSE	PROTO-ONCOGENE TYROSIN	4.53e-04
41	94	23.6	1130	1 ABL1_HUMAN	PROTO-ONCOGENE TYROSIN	4.53e-04
42	94	23.6	1182	1 ABL2_HUMAN	TYROSINE-PROTEIN KINAS	4.53e-04
43	94	23.6	1293	1 ENTF_ECOLI	ENTEROBACTIN SYNTHETAS	4.53e-04
44	93	23.3	539	1 YES_CANFA	PROTO-ONCOGENE TYROSIN	7.02e-04
45	91	22.8	508	1 LCK_HUMAN	PROTO-ONCOGENE TYROSIN	1.67e-03

ALIGNMENTS

RESULT	ID	SRM_MOUSE	STANDARD;	PRT;	496 AA.
AC	062270	0622360			
DT	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	DE TYROSINE-PROTEIN KINASE SRM (EC 2.7.1.112) (PTK70).				
GN	SRMS OR SRM.				
OS	MUS MUSCULUS (MOUSE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6; TISSUE=THYMUS;				
RX	MEDLINE; 97369678				
RA	KAWACHI Y., NAKAUCHI H., OTSUKA F.;				
RT	"Isolation of a cDNA encoding a tyrosine kinase expressed in murine skin";				
RL	EXP. DERMATOL. 21:533-538(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=LUNG;				
RX	MEDLINE; 95021220.				
RA	KOHMURA N., YAGI T., TOMOOKA Y., OYANAGI M., KOMINAMI R., TAKEDA N.,				
RA	CHIBA J., IKAWA Y., AIZAWA S.;				
RT	"A novel nonreceptor tyrosine kinase, Src: cloning and targeted disruption.";				
RL	MOL. CELL. BIOL. 14:6915-6925(1994).				
CC	-!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.				
CC	-!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC DOMAIN. BELONGS TO THE SRC SUBFAMILY.				
CC	-!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.				
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; D49427; G684972; -				
DR	EMBL; D26186; G529073; -				
DR	MGD; MGI:101865; SRMS.				
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.				

```
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; PKINASE; 1.
DR HSP; P1362; IFGI.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 55 116 SH3.
FT DOMAIN 124 216 SH2.
FT DOMAIN 234 495 PROTEIN KINASE.
FT NP_BIND 240 248 ATP (BY SIMILARITY).
FT BINDING 262 262 ATP (BY SIMILARITY).
FT ACT_SITE 354 354 BY SIMILARITY.
FT MOD_RES 384 384 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 78 78 R -> G (IN REF. 2).
FT CONFLICT 236 238 LRK -> FGR (IN REF. 2).
FT CONFLICT 278 278 N -> I (IN REF. 2).
SQ SEQUENCE 496 AA; 55731 MW; FD44DEF6 CRC32;

Query Match 73.2%; Score 292; DB 1; Length 496;
Best Local Similarity 74.1%; Pred. No. 1.15e-51;
Matches 40; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 60 FLALYDFTARCAEELSVGRDLYALKEGDYIFQRLSGPSTGLVPVTVLAK 113
| ||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 56 FLALYDFTARCGGELSVGRDRLCALEEGGYIFARRLSGQPSAGLVPITHAK 109

RESULT 2
ID STK_HYDAT STANDARD; PRT; 509 AA.
AC P1713;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE STK (EC 2.7.1.112) (P57-STK).
GN STK.
OS HYDRA ATTENUATA (HYDRA) (HYDRA VULGARIS).
OC EUKARYOTA; METAZOA; CNIDARIA; HYDROZOA; HYDROIDA; ANTHOMEDUSAE;
OC HYDRIAE; HYDRA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90066418.
RA BOSCH T.C.G., UNGER T.F., FISHER D.A., STEELE R.E.;
RT "Structure and expression of STK, a src-related gene in the simple
metazoan Hydra attenuata."
RL MOL. CELL. BIOL. 9:4141-4151(1989).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M25245; G159274;
CC PIR; A34094; TVHAST.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
```

```
DR PFAM; PF00069; pkinese; 1.
DR HSP; P00523; IPRL.
KW TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; PHOSPHORYLATION;
KW TRANSFERASE; ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
FT DOMAIN 59 120 SH3.
FT DOMAIN 126 218 SH2.
FT DOMAIN 240 495 PROTEIN KINASE.
FT NP_BIND 246 254 ATP (BY SIMILARITY).
FT BINDING 268 268 ATP (BY SIMILARITY).
FT ACT_SITE 360 360 BY SIMILARITY.
FT MOD_RES 390 390 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 509 AA; 56885 MW; 2B724CE9 CRC32;

Query Match 33.6%; Score 134; DB 1; Length 509;
Best Local Similarity 40.7%; Pred. No. 2.07e-12;
Matches 22; Conservative 11; Mismatches 19; Indels 2; Gaps 2;

Db 64 FVALDYEARISEDLSFKKGERLQIINTADGDMWYARSLTN-SEGYPSTYVA 116
|:||||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
QY 56 FLALYDFTARCGGELSVRRGRLCALE-EGGYIFARRLSGQPSAGLVPITHVA 108

RESULT 3
ID SRC2_XENLA STANDARD; PRT; 531 AA.
AC P13116;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRC-2 (EC 2.7.1.112) (P60-SRC-2).
GN SRC-2.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPOIDEA; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89278134.
RA STEELE R.E., UNGER T.F., MARDIS M.J., FERO J.B.;
RT "The two Xenopus laevis SRC genes are co-expressed and each produces
functional p60src."
RL J. BIOL. CHEM. 264:10649-10653(1989).
RN [2]
RP SEQUENCE OF 438-491 FROM N.A.
RX TISSUE-ERYTHROCYTE;
RX MEDLINE; 85215578.
RA STEELE R.E.;
RT "Two divergent cellular src genes are expressed in Xenopus laevis."
RL NUCLEIC ACIDS RES. 13:1747-1761(1985).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M23422; G214797;
CC EMBL; M30858; G555569;
CC PIR; B34104; B34104.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
```

[illegible]

DR HSP: P06239; 3LCK.  
KW PROTO-ONCOGENE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; TRANSFERASE;  
KW ATP-BINDING; MYRISTYLATION; SH2 DOMAIN; SH3 DOMAIN; PALMITATE;  
KW LIPOPROTEIN.  
FT INIT\_MET 0 0 PROBABLE.  
FT LIPID 1 1 MYRISTATE (BY SIMILARITY).  
FT LIPID 2 2 PALMITATE (BY SIMILARITY).  
FT LIPID 4 4 PALMITATE (BY SIMILARITY).  
FT LIPID 59 119 SH3.  
FT DOMAIN 125 222 SH2.  
FT DOMAIN 243 496 PROTEIN KINASE.  
FT NP\_BIND 249 257 ATP (BY SIMILARITY).  
FT BINDING 271 271 ATP (BY SIMILARITY).  
FT ACT\_SITE 362 362 BY SIMILARITY.  
FT MOD\_RES 392 392 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 503 503 PHOSPHORYLATION.  
SQ SEQUENCE 507 AA; 58008 MW; 06C5A291 CRC32;  
Query Match 29.8%; Score 119; DB 1; Length 507;  
Best Local Similarity 35.8%; Pred. No. 3.85e-09;  
Matches 19; Conservative 15; Mismatches 18; Indels 1; Gaps 1;  
Db 65 VALYDYEPTDGLGKQEKRLVLEESGEWRAQSLTTGQ-EGLPHPNFVAM 116  
QY 57 LALYDFTARCAGGELSVRRGDRLCALAEEGGYIFARRLSGQPSAGLVPIITHVAK.109..  
RESULT 6 STANDARD; PRT; 536 AA.  
ID FYN\_XIPHE  
AC P27446;  
DT 01-AUG-1992 (REL. 23, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).  
GN FYN.  
OS XIPHOPHORUS HELLERI.  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;  
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA;  
OC CYPRINODONTIFORMES; CYPRINODONTOIDEI; POECILIIDAE; XIPHOPHORUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RIO LANCETILLA;  
RX MEDLINE; 91187435.  
RA HANNIG G., OTTILIE S., SCHARTL M.;  
RT "Conservation of structure and expression of the c-yes and fyn genes  
in lower vertebrates.";  
RL ONCOGENE 6:361-369(1991).  
CC -!- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF  
PHOSPHATIDYLINOSITOL 3-KINASE.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; X54971; G64482;  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50001; SH2; 1.  
DR PROSITE; PS50002; SH3; 1.  
DR PFAM; PF00017; SH2; 1.  
DR PFAM; PF00018; SH3; 1.

DR PFAM; PF00069; pkinase; 1.  
DR HSP; P06241; 1AOT.  
KW PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;  
KW ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;  
KW LIPOPROTEIN.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT LIPID 1 1 MYRISTATE (BY SIMILARITY).  
FT LIPID 2 2 PALMITATE (BY SIMILARITY).  
FT LIPID 5 5 PALMITATE (BY SIMILARITY).  
FT LIPID 81 142 SH3.  
FT DOMAIN 148 245 SH2.  
FT DOMAIN 270 523 PROTEIN KINASE.  
FT MOD\_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
FT NP\_BIND 276 284 ATP (BY SIMILARITY).  
FT BINDING 298 298 ATP (BY SIMILARITY).  
FT ACT\_SITE 389 389 BY SIMILARITY.  
FT MOD\_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 530 530 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 536 AA; 60316 MW; BA03DB12 CRC32;  
Query Match 29.6%; Score 118; DB 1; Length 536;  
Best Local Similarity 38.9%; Pred. No. 6.29e-09;  
Matches 21; Conservative 11; Mismatches 20; Indels 2; Gaps 2;  
Db 86 FVALYDYEPTDGLGKQEKRLVLEESGEWRAQSLTTGQ-EGSGYIPSNVA 138  
QY 56 FLALYDFTARCAGGELSVRRGDRLCALAEEGGYIF-ARRLSGQPSAGLVPIITHVA 108  
RESULT 7  
ID FRK\_HUMAN  
AC P42685; Q13128;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE FRK (EC 2.7.1.112) (NUCLEAR TYROSINE PROTEIN  
DE KINASE RAK).  
GN FRK.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LYMPHOID;  
RX MEDLINE; 94171047.  
RA LEE J., WANG Z., LUOH S.-M., WOOD W.I., SCADDEN D.T.;  
RT "Cloning of FRK, a novel human intracellular SRC-like tyrosine  
kinase-encoding gene.";  
RL GENE 138:247-251(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95210168.  
RA CANOE W.G., CRAVEN R.J., BERGMAN M., XU L.H., ALITALO K., LIU E.T.;  
RT "Rak, a novel nuclear tyrosine kinase expressed in epithelial cells.";  
RL CELL GROWTH DIFFER. 5:1347-1355(1994).  
RN [3]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE; 93293373.  
RA CANOE W.G., CRAVEN R.J., WEINER T.M., LIU E.T.;  
RT "Novel protein kinases expressed in human breast cancer.";  
RL INT. J. CANCER 54:571-577(1993).  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -!- TISSUE SPECIFICITY: RESTRICTED TO CELLS LINES DERIVED FROM TISSUES  
OF LYMPHOID, BRAIN, BREAST, COLON AND BLADDER ORIGIN.  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -









FT	DOMAIN	267	520		PROTEIN KINASE.
FT	N_BIND	273	281		ATP (BY SIMILARITY).
FT	BINDING	295	295		ATP (BY SIMILARITY).
FT	C_SITE	386	386		BY SIMILARITY.
FT	MOD_RES	416	416		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ	SEQUENCE	568 AA;	63632 MW; 7F080D52 CRC32;		

Query Match 27.8%; Score 111; DB 1; Length 568;  
Best Local Similarity 33.3%; Pred. No. 1.85e-07;  
Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

Dd	86	FVALYDYESTETDLSPFKGERLQIVNTGDTWLAHSLTGTQT-GVIPSNYVA	138
Qy	56	FLALYDFARGCGELSVRRGRDLCALEEG-GGYIFARRLSQGQSAGLVPIITHA	108

```

|:||||: |: || :||| :: | :|: | : | : ||
|:||||: |: || :||| :: | :|: | : | : ||
|:||||: |: || :||| :: | :|: | : | : ||
|:||||: |: || :||| :: | :|: | : | : ||

```

RESULT 13  
ID SRC\_AVIS2 STANDARD; PRT; 597 AA.  
AC P15054;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE TRANSFORMING PROTEIN SRC (EC 2.7.1.112) (P60-SRC).  
GN V-SRC.  
OS AVIAN SARCOMA VIRUS (STRAIN PR2257).  
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89094972.  
RA GERYK J., DEZELEE P., BARNIER J.V., SVOBODA J., NEHYBA J., KARAKOZ I., RYNDITCH A.V., YATSULA B.A., CALOTHY G.; "transduction of the cellular src gene and 3' adjacent sequences in avian sarcoma virus pcr2257";  
RT J. VIROL. 63:481-492(1989).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP RA YATSULA B.A., GERYK J., SVOBODA J., RYNDITCH A.V., CALOTHY G., DEZELEE P.;  
RL SUBMITTED (APR-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.  
[3]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 146-249.  
RA HOLLAND D.R., LUNNEY E.A., PLUMMER M.S., MUELLER W.T., MCCONNELL P., RA PAVLOVSKY A., PARA K.S., SHAHRIPOUR A., HUMBLETT C., SAWYER T.K., RUBIN J.R.;  
RL SUBMITTED (MAY-1997) TO THE PDB DATA BANK  
CC -!- FUNCTION: THIS PHOSPHOPROTEIN, REQUIRED FOR BOTH THE INITIATION AND THE MAINTENANCE OF NEOPLASTIC TRANSFORMATION, IS A PROTEIN KINASE THAT CATALYZES THE PHOSPHORYLATION OF TYROSINE RESIDUES IN VITRO.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
-----  
CC EMBL; M21526; G210265; -  
CC EMBL; X51863; G394713; -  
DR PIR; A30174; TVFVPR.  
DR PDB; 1BK1; 23-JUL-97.  
DR PDB; 1BKM; 07-JUL-97.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

```
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR TYROSINE-PROTEIN KINASE; TRANSFORMING PROTEIN; ONCOGENE;
KW TRANSFERASE; PHOSPHORYLATION; ATP-BINDING; MYRISTYLATION;
KW SH3 DOMAIN; SH2 DOMAIN; 3D-STRUCTURE.
FT LIPID 2 2 MYRISTATE.
FT DOMAIN 81 142 SH3.
FT DOMAIN 148 245 SH2.
FT DOMAIN 267 520 PROTEIN KINASE.
FT NP_BIND 273 281 ATP (BY SIMILARITY).
FT BINDING 295 295 ATP (BY SIMILARITY).
FT ACT_SITE 366 366 BY SIMILARITY.
FT MOD_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 587 AA; 65800 MW; 88DFD703 CRC32;

Query Match 27.8%; Score 111; DB 1; Length 587;
Best Local Similarity 33.3%; Pred. No. 1.85e-07;
Matches 18; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

Db 86 FVALDYESTRTDLSFKKGERLQIVNTEGDWLAHSLTGTQ-TGYPSNYVA 138
1:1111:1 :11:1111: : : : : : : : : : : : : : : : : :
QY 56 FLALDYFTARCGELSVRRGDRLCALLEG-GGYIFARLSGQPSAGLVPTHTVA 108
1:1111:1 :11:1111: : : : : : : : : : : : : : : : : :

RESULT 14
ID YRK_CHICK STANDARD; PRT; 535 AA.
AC Q02977.
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE YRK (EC 2.7.1.112) (P60-YRK)
DE (YES RELATED KINASE).
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEIGHORN; TISSUE-BRAIN, AND KIDNEY;
RX MEDLINE; 93205395.
RA SUDOL M., GREULICH H., NEWMAN L., SARKAR A., SUKAGAWA J.,
RA YAMAMOTO T.;
RT "A novel yes-related kinase, Yrk, is expressed at elevated levels in
neural and hematopoietic tissues."
RL ONCOGENE 8:823-831(1993).
CC -!- FUNCTION: MAY PARTICIPATE IN SIGNALING PATHWAYS.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
PROTEIN TYROSINE PHOSPHATE.
CC -!- TISSUE SPECIFICITY: THERE ARE ELEVATED LEVELS OF THIS PROTEIN IN
NEURAL AND HEMATOPOIETIC TISSUES.
CC -!- PTM: PHOSPHORYLATED.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X67786; G63896;
DR PIR; S29553; S29553.
DR PIR; S29626; S29626.
DR PIR; S33569; S33569.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P06241; IAO.
KW PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;
KW LIPOPROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
FT LIPID 1 1 MYRISTATE (BY SIMILARITY).
FT LIPID 2 2 PALMITATE (BY SIMILARITY).
FT LIPID 5 5 PALMITATE (BY SIMILARITY).
FT DOMAIN 80 141 SH3.
FT DOMAIN 147 244 SH2.
FT DOMAIN 269 522 PROTEIN KINASE.
FT NP_BIND 275 283 ATP (BY SIMILARITY).
FT BINDING 297 297 ATP (BY SIMILARITY).
FT ACT_SITE 388 388 BY SIMILARITY.
FT MOD_RES 418 418 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 529 529 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 535 AA; 59871 MW; 50B7CFF1 CRC32;

Query Match 27.6%; Score 110; DB 1; Length 535;
Best Local Similarity 35.2%; Pred. No. 2.98e-07;
Matches 19; Conservative 15; Mismatches 18; Indels 2; Gaps 2;

Db 85 FIALYDEARTEDDLSFKGKFKHIINTEGDWEARSLSS-GATGYTPSNVA 137
1:1111:1 :11:1111: : : : : : : : : : : : : : : : : :
QY 56 FLALDYFTARCGELSVRRGDRLCALLEGGYIF-ARRLSGQPSAGLVPTHTVA 108
1:1111:1 :11:1111: : : : : : : : : : : : : : : : : :

RESULT 15
ID YES_XIPHE STANDARD; PRT; 544 AA.
AC P27447.
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE YES (EC 2.7.1.112) (P61-YES)
DE (C-YES).
OS YES.
OS XIPHOPHORUS HELLERI.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA.
OC CYPRINODONTIFORMES; CYPRINODONTIOIDEI; POECILIIDAE; XIPHOPHORUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIO LANCETILLA;
RX MEDLINE; 91187435.
RA HANNIG G., OTTILIE S., SCHARTL M.;
RT "Conservation of structure and expression of the c-yes and fyn genes
in lower vertebrates."
RL ONCOGENE 6:361-369(1991).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X54970; G64484;
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
```



Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	127	31.8	517	5	Q94879	DSRC41	3.36e-09
2	115	28.8	526	14	Q07461	TYROSINE-PROTEIN KINAS	6.32e-07
3	115	28.8	527	13	Q91952	C-SRC TYROSINE KINASE.	6.32e-07
4	113	28.3	512	11	Q61745	B-CELL SRC-HOMOLOGY TY	1.48e-06
5	113	28.3	512	11	Q61364	B-CELL SRC-HOMOLOGY TY	1.48e-06
6	111	27.8	193	13	Q90993	NON-TYROSIN PROTEIN KI	3.47e-06
7	111	27.8	496	5	Q45539	F49B2.5 PROTEIN.	3.47e-06
8	111	27.8	517	11	Q63206	FGF MRNA.	3.47e-06
9	111	27.8	533	13	Q90992	C-SRC.	3.47e-06
10	111	27.8	533	13	Q98915	GENE C-SRC PRODUCING P	3.47e-06
11	111	27.8	587	14	Q64817	PROTEIN-TYROSINE KINAS	3.47e-06
12	109	27.3	517	5	Q70500	SRC-TYPE PROTEIN TYROS	8.04e-06
13	109	27.3	833	11	Q35177	NEURAL PRECURSOR CELL	8.04e-06
14	109	27.3	834	4	Q14511	ENHANCER OF FILAMENTATI	8.04e-06
15	107	26.8	517	11	Q61404	GARDNER-RASHEED FELINE	1.85e-05
16	106	26.6	534	4	Q16248	P59FYN.	2.81e-05
17	106	26.6	537	11	Q82844	PROTO-ONCOGENE FYN.	2.81e-05
18	105	26.3	506	11	Q82662	SRC RELATED TYROSINE K	4.25e-05
19	104	26.1	204	14	Q95730	PP60-SRC PROTEIN (FRAG	6.41e-05
20	104	26.1	285	14	Q85476	(RECOVERED INSERTION M	6.41e-05



[illegible]



RESULT 13

Search completed: Thu May 20 12:49:25 1999  
Job time : 179 secs.

\*\*\*\*\*

W P E L H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:50:50 1999; Maspar time 11.20 Seconds  
Tabular output not generated. 176.653 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (120-212) from US09099053.pap (4 of 6)  
Perfect Score: 699  
Sequence: 1 WFGVSVRTQAQQLLLSPN.....LTYKANKWKLQNPLOPC 93

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 29.030; Variance 115.410; scale 0.252

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	699	100.0	488 39	W89248	Human SAD	8.52e-61
2	260	37.2	298 15	R84193	Megakaryocyte kinase	1.62e-15
3	260	37.2	505 14	R85929	Protein tyrosine-kinase	1.62e-15
4	260	37.2	505 8	R41941	prk gene Lprk-2 prod.	1.62e-15
5	241	34.5	101 26	W31184	Human p56-lck protein	1.20e-13
6	241	34.5	134 18	W02120	DETI-DET2-spacer-ek-1	1.20e-13
7	241	34.5	134 24	W19624	Human lck SH2 domain	1.20e-13
8	241	34.5	134 21	W03982	DETI-DET2-spacer-ek-1	1.20e-13
9	241	34.5	134 23	W11286	DETI-DET2-spacer-ek-1	1.20e-13
10	241	34.5	224 21	W14788	FKBP-LCK:SH2 fusion p	1.20e-13
11	240	34.3	451 12	R63088	Breast tumor kinase	1.51e-13
12	239	34.2	98 33	W59760	Amino acid sequence o	1.89e-13
13	235	33.6	98 33	W59756	Amino acid sequence o	4.85e-13
14	235	33.6	533 8	R39705	Chicken pp60 c-src pr	4.85e-13
15	234	33.5	98 33	W59763	Amino acid sequence o	5.83e-13
16	233	33.3	330 37	W76830	Human GRBP protein.	7.30e-13

17	230	32.9	98 33	W59757	Amino acid sequence o	1.43e-12
18	229	32.8	102 13	R72090	Lck SH2 region.	1.79e-12
19	227	32.5	98 33	W59758	Amino acid sequence o	2.81e-12
20	225	32.2	130 24	W19623	Human src SH2 domain	4.40e-12
21	225	32.2	130 21	W03981	DETI-DET2-spacer-ek-s	4.40e-12
22	225	32.2	130 23	W11285	DETI-DET2-spacer-ek-s	4.40e-12
23	225	32.2	130 18	W02119	DETI-DET2-spacer-ek-s	4.40e-12
24	225	32.2	536 8	R39706	Human pp60 C-src prot	4.40e-12
25	224	32.0	105 13	R72088	Src SH2 region.	5.50e-12
26	222	31.8	417 3	R14201	(Beta-galactosidase N	8.61e-12
27	218	31.2	1182 3	R15157	Abelson Related Gene,	2.10e-11
28	217	31.0	98 33	W59759	Amino acid sequence o	2.63e-11
29	217	31.0	106 21	W03989	SH2 domain from human	2.63e-11
30	217	31.0	106 18	W02127	Human Grb2 SH2 domain	2.63e-11
31	217	31.0	217 21	W14004	Human GRB2.	2.63e-11
32	217	31.0	217 29	W02070	Growth factor recepto	2.63e-11
33	217	31.0	217 16	R85918	Human GRB-2.	2.63e-11
34	215	30.8	1146 3	R15156	Abelson Related Gene,	4.11e-11
35	210	30.0	217 15	R84636	Grb2 protein.	1.25e-10
36	208	29.8	217 24	W18063	Growth factor recepto	1.95e-10
37	205	29.3	1290 15	R30583	Phospholipase C-gamma	3.79e-10
38	204	29.2	384 21	W14787	FKBP-SYK:SH2 fusion p	4.73e-10
39	204	29.2	612 20	W09321	Human mast cell-deriv	4.73e-10
40	204	29.2	630 12	R62688	Tyrosine-kinase Syk.	4.73e-10
41	202	28.9	99 13	R72089	Abi SH2 region.	7.36e-10
42	198	28.3	117 18	W02124	Human fyn SH2 domain	1.78e-09
43	198	28.3	117 21	W03986	SH2 domain from human	1.78e-09
44	193	27.6	112 21	W03988	SH2 domain from human	5.34e-09
45	193	27.6	593 10	R52991	Human protein-tyrosin	5.34e-09

## ALIGNMENTS

RESULT 1  
ID W89248 standard; Protein; 488 AA.  
AC W89248:  
DT 10-MAR-1999 (first entry)  
DE Human SAD.  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease.  
OS Homo sapiens.  
PN W09849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGE-) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Orust S, Peles E, Plowman GD;  
DR WPI; 99-009434/01.  
DR N-PSDB: V81743.  
PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Claim 2; Page 154-155; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC present sequence represents human SAD.  
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
CC substances that modulate their activity (i.e. agonists and antagonists,  
CC including NPP) in vivo or in vitro. These substances are used to treat  
CC or prevent diseases associated with abnormal signal transduction  
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
CC serine/threonine kinase) are used to promote neuronal survival,  
CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
CC diseases. Nucleic acid fragments of the polynucleotides encoding the

RESULT 3  
ID R85929 standard; Protein; 505 AA.  
AC R85929;

PT are of human mega-karyocytic origin  
PS. Claim 3; Fig 5; 6Opp; English.  
CC The pTKs were identified using two sets of degenerative  
CC oligonucleotide primers: a first set which amplifies all pTK DNA  
CC segments (Q49743-44), and a second set which amplifies highly  
CC conserved sequences present in the catalytic domain of the c-kit  
CC subgroup of pTKs (Q49745-46). The pTK genes identified are described  
CC in Q49747-57 and R41897-02.  
CC The LpTKs are expressed in lymphocytic cells, as well as  
CC megakaryocytic cells. The partial and full-length LpTK2 gene  
CC sequences are given in Q49749 and Q49754 respectively. The  
CC protein sequence corresp. to Q49749 is claimed (Claim 7) and  
CC stated as given in the specification, however is missing from  
CC the publication.  
SQ Sequence 505 AA;





FT /note= "similar to PTK SN3 domain"

CC the method of the invention to inactivate signal transduction proteins  
 CC in the treatment of e.g. cancer, or for targeting specific SH2 domains  
 CC for diagnostics. The modified Src SH2 domain is useful as a research  
 CC tool to study signal transduction in general and for a variety of  
 CC applications such as diagnostics, therapeutics, and drug design. It  
 CC detects aberrant phosphorylation events including those associated with  
 CC tumour progression of breast and ovarian carcinomas for tumour  
 CC diagnosis and staging. It interferes with signal transduction pathways  
 CC to treat disorders resulting from aberrant signal transduction pathways  
 CC such as cancer, autoimmune disease, and allergies. Interference with  
 CC the signal generated by the signal transduction protein prevents  
 CC further transmission of the signal to other components of the cell. It  
 CC also binds to an aberrant signal transduction protein thereby blocking  
 CC the binding of the wild type SH2 domains, or inactivates the  
 CC phosphorylated signal transduction protein by removing a phosphate group.  
 CC This sequence is not given in the specification, but is created using  
 CC W59756, and information provided.  
 CC Sequence 98 AA;

Query Match 34.2%; Score 239; DB 33; Length 98;  
 Best Local Similarity 35.7%; Pred. No. 1.89e-13;  
 Matches 35; Conservative 28; Mismatches 30; Indels 5; Gaps 3;

Db 1 wyfgkitreserlllnpenprgtflvresettkgayclsvsdfdnakglnvkhkirk1 60  
 QY 120 WYFGSVSRTOAQLLSPNPEGAFILRPSSESLGYSLSVRA-Q-AK---VCHYRVMSA 174  
 Db 61 dsggfyitsrtqfssllqqlvayv 98  
 QY 175 ADGSLYLOKGRFLPGLELLTYKANWKLQNPLQPC 212

RESULT 13  
 ID W59756 standard; peptide; 98 AA.  
 AC W59756;  
 DT 12-OCT-1998 (first entry)  
 DE Amino acid sequence of the SH2 domain src.  
 KW SH2 domain; signal transduction protein; cancer; SH2 domain src; tumour;  
 KW phosphorylation; breast cancer; ovarian cancer; autoimmune disease;  
 KW allergies.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Domain 1..98  
 FT /note= "SH2 domain src"

US5786454-A.  
 28-JUL-1998.  
 PF 16-SEP-1994; 308086.  
 PR 16-SEP-1994; US-308086.  
 PA (UNIW ) UNIV WASHINGTON SCHOOL MED.  
 PI Shaw A, Waksman G;  
 DR WPI; 98-436608/37.  
 PT Modified SH2 domains of intracellular proteins - useful for  
 PT inactivating signal transduction proteins in the treatment of e.g.  
 PT cancer, or for targeting specific SH2 domains for diagnostics  
 PS Disclosure; Fig 2; 22pp; English.  
 CC This is the amino acid sequence of the SH2 domain src, modified in the  
 CC method of the invention where modified SH2 domains of intracellular  
 CC proteins are used for inactivating signal transduction proteins in the  
 CC treatment of e.g. cancer, or for targeting specific SH2 domains for  
 CC diagnostics. The modified Src SH2 domain is useful as a research tool  
 CC to study signal transduction in general and for a variety of  
 CC applications such as diagnostics, therapeutics, and drug design. It  
 CC detects aberrant phosphorylation events including those associated with  
 CC tumour progression of breast and ovarian carcinomas for tumour  
 CC diagnosis and staging. It interferes with signal transduction pathways  
 CC to treat disorders resulting from aberrant signal transduction pathways  
 CC such as cancer, autoimmune disease, and allergies. Interference with  
 CC the signal generated by the signal transduction protein prevents  
 CC further transmission of the signal to other components of the cell. It  
 CC also binds to an aberrant signal transduction protein thereby blocking  
 CC the binding of the wild type SH2 domains, or inactivates the  
 CC phosphorylated signal transduction protein by removing a phosphate group.  
 CC Sequence 98 AA;

Query Match 33.6%; Score 235; DB 33; Length 98;  
 Best Local Similarity 38.6%; Pred. No. 4.65e-13;  
 Matches 32; Conservative 26; Mismatches 20; Indels 5; Gaps 3;

Db 1 wyfgkitreserlllnpenprgtflvresettkgayclsvsdfdnakglnvkhkirk1 60  
 QY 120 WYFGSVSRTOAQLLSPNPEGAFILRPSSESLGYSLSVRA-Q-AK---VCHYRVMSA 174  
 Db 61 dsggfyitsrtqfssllqqlvayv 83  
 QY 175 ADGSLYLOKGRFLPGLELLTY 197

RESULT 14  
 ID R39705 standard; Protein; 533 AA.  
 AC R39705;  
 DT 23-DEC-1993 (first entry)  
 DE Chicken pp60 c-src protein.  
 KW Endothelial; tyrosine kinase protein; pp60 c-src; ss.  
 OS Gallus gallus.  
 PN W09314193-A.  
 PD 22-JUL-1993.  
 PF 05-JAN-1993; US00445.  
 PR 06-JAN-1992; US-820011.  
 PA (UYVA ) UNIV YALE.  
 PI Bell L, Luthringer DJ, Madri JA, Warren SL;  
 DR P-PSDB; R39705.  
 PT Genetically engineered endothelial cells - which exhibit enhanced  
 PT cell migration, urokinase-type plasminogen activator activity,  
 PT and reduced mononuclear cell adhesion and fibronectin prodn  
 PS Disclosure; Page 64-66; 91pp; English.  
 CC The DNA encoding a portion or (more preferably) the entire pp60  
 CC c-src polypeptide (Given in Q46687) is used to transform endothelial  
 CC cells. Transformed cells produce increased amounts of pp60 c-src and  
 CC have improved therapeutic properties. They migrate at faster rates  
 CC than non-transformed counterparts; have an enhanced ability to  
 CC inhibit the formation of thrombi and/or dissolve thrombi once they  
 CC have formed and exhibit reduced mononuclear cell adhesion. They can  
 CC also be used to improve the success of surgical procedures such as  
 CC coronary angioplasty, heart bypass surgery, vessel graft and stent  
 CC implantation.  
 CC Sequence 533 AA;

Query Match 33.6%; Score 235; DB 8; Length 533;  
 Best Local Similarity 38.6%; Pred. No. 4.65e-13;  
 Matches 32; Conservative 26; Mismatches 20; Indels 5; Gaps 3;

Db 148 wyfgkitreserlllnpenprgtflvresettkgayclsvsdfdnakglnvkhkirk1 207  
 QY 120 WYFGSVSRTOAQLLSPNPEGAFILRPSSESLGYSLSVRA-Q-AK---VCHYRVMSA 174  
 Db 208 dsggfyitsrtqfssllqqlvayv 230  
 QY 175 ADGSLYLOKGRFLPGLELLTY 197

RESULT 15  
 ID W59763 standard; peptide; 98 AA.  
 AC W59763;  
 DT 12-OCT-1998 (first entry)  
 DE Amino acid sequence of the SH2 domain src mutant M3.  
 KW SH2 domain; signal transduction protein; cancer; SH2 domain src; tumour;  
 KW phosphorylation; breast cancer; ovarian cancer; autoimmune disease;  
 KW allergies.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Domain 1..98  
 FT /note= "SH2 domain src"  
 FT US5786454-A.  
 PN 28-JUL-1998.  
 PF 16-SEP-1994; 308086.



PR 16-SEP-1994: US-308086.  
 PA (UNIW ) UNIV WASHINGTON SCHOOL MED.  
 PI Shaw A, waksman G;  
 DR WPI: 98-436608/37.  
 PT Modified SH2 domains of intracellular proteins - useful for  
 PT inactivating signal transduction proteins in the treatment of e.g.  
 PT cancer, or for targeting specific SH2 domains for diagnostics  
 PS Example 2; Column : 22pp; English.  
 CC This is the amino acid sequence of a SH2 domain src mutant M2, used in  
 CC the method of the invention to inactivate signal transduction proteins  
 CC in the treatment of e.g. cancer, or for targeting specific SH2 domains  
 CC for diagnostics. The modified Src SH2 domain is useful as a research  
 CC tool to study signal transduction in general and for a variety of  
 CC applications such as diagnostics, therapeutics, and drug design. It  
 CC detects aberrant phosphorylation events including those associated with  
 CC tumour progression of breast and ovarian carcinomas for tumour  
 CC diagnosis and staging. It interferes with signal transduction pathways  
 CC to treat disorders resulting from aberrant signal transduction pathways  
 CC such as cancer, autoimmune disease, and allergies. Interference with  
 CC the signal generated by the signal transduction protein prevents  
 CC further transmission of the signal to other components of the cell. It  
 CC also binds to an aberrant signal transduction protein thereby blocking  
 CC the binding of the wild type SH2 domains, or inactivates the  
 CC phosphorylated signal transduction protein by removing a phosphate group.  
 CC This sequence is not given in the specification, but is created using  
 CC W59756, and information provided.  
 SQ Sequence 98 AA;

Query Match 33.5%; Score 234; DB 33; Length 98;  
 Best Local Similarity 38.6%; Pred. No. 5.83e-13;  
 Matches 32; Conservative 26; Mismatches 20; Indels 5; Gaps 3;  
 Db 1 wyfgkitrreserlllnpenprgtflvresettkgayclsvsdfdnakgnvkhkykirk1 60  
 QY 120 WYFGSVSRTOAQQLLLSPNPGAFLLIRPSESLGGYSLSVRA-Q-AK---VCHRYVSMA 174  
 Db 61 dsqggyfsrtqfsslqlvayy 83  
 QY 175 ADGSLYLQKGRLPFGLELLTY 197

Search completed: Thu May 20 12:53:20 1999  
 Job time : 150 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```
Run on: Thu May 20 12:53:40 1999; MasPar time 9.55 Seconds
364.544 Million cell updates/sec
Tabular output not generated.
```

LLSPPN.....LLTYKANWKLI ONPLLOPC 93

Gap 11

Post-processing: Minimum Match 0%  
Listing first 45

Statistics: Mean 39.723; Variance 76.072; scale 0.522

## SUMMARIES

Result No.	Query		DB	ID	Description	Pred. No.
	Match	Score				
1	625	89.4	496	2	I56322	8.61e-106
2	625	89.4	496	2	A56040	8.61e-106
3	276	39.5	512	2	I49552	1.29e-33
4	269	38.5	506	2	S24553	3.01e-31
5	262	37.5	505	2	S24550	6.94e-31
6	260	37.2	505	2	I38396	1.70e-30
7	241	34.5	509	1	OKHULK	7.83e-27
8	240	34.3	451	2	S49016	1.22e-26
9	238	34.0	509	2	I48845	2.94e-26
10	238	34.0	509	2	A23639	2.94e-26
11	235	33.6	536	1	TVFV60	1.10e-25
12	235	33.6	533	1	TVFCHS	1.10e-25
13	235	33.6	557	1	TVFV52	1.10e-25
14	235	33.6	587	1	TVFV52	1.10e-25
15	233	33.3	505	1	TVFHUC	1.10e-25
16	233	33.3	544	2	I51593	2.64e-25
17	232	33.2	509	1	TVFV51	2.64e-25
18	232	33.2	568	1	TVHAST	4.09e-25
19	230	32.9	499	2	A40092	9.81e-25
20	230	32.9	503	2	J01321	9.81e-25
21	229	32.8	523	1	TVFVMT	9.81e-25
22	229	32.8	545	2	S52313	1.52e-24
23	229	32.8	546	2	S52314	1.52e-24

```

#authors Kohmura, N.; Yagi, T.; Tomooka, Y.; Oyanagi, M.; Kominami,
#journal R.; Takeda, N.; Chiba, J.; Ikawa, Y.; Aizawa, S.
#title Mol. Cell. Biol. (1994) 14:6915-6925
#title A novel nonreceptor tyrosine kinase, Srm: cloning and
#title targeted disruption.
#accession A56040
#status preliminary
#molecule_type mRNA
#residues 1-496 ##label KOH
#cross-references GB:D26186; NID:g529072; PID:d1005873; PID:g529073
GENETICS
#map_position 2
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
#keywords kinases; protein kinase homology; SH3 homology
#feature ATP; phosphotransferase
FEATURE
62-111 #domain SH3 homology #label SH3\
232-491 #domain protein kinase homology #label KIN\
240-248 #region protein kinase ATP-binding motif
SUMMARY #length 496 #molecular-weight 55593 #checks 301
Query Match 89.4%; Score 625; DB 2; Length 496;
Best Local Similarity 82.8%; Pred. No. 8.61e-106;
Matches 77; Conservative 11; Mismatches 5; Indels 0; Gaps 0;
Db 124 WYFGISRAQAQQLLSPANAPGAFILRPSESSIGGYSLSVRAQAKVCHYRICMAPSGSL 183
QY 120 WYFGSVSRTOAQLLSPNPEGAFILRPSESSIGGYSLSVRAQAKVCHYRVSMAADGSL 179
Db 184 YLQEQQLPFLDALLAYKTNKWLQNPLQPC 216
QY 180 YLQKGRLPFGLELLTYKANKWLQNPLQPC 212
RESULT 3
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) bsk/iyk - mouse
ALTERNATE_NAMES intestinal tyrosine kinase
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
09-Apr-1998
ACCESSIONS I49552 #type complete
REFERENCE I49552
#authors Oberg-Welsh, C.; Welsh, M.
#journal Gene (1995) 152:239-242
#title Cloning of BSK, a murine FRK homologue with a specific
#title pattern of tissue distribution.
#cross-references MUID:95137395
#accession I49552
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-512 ##label RES
#cross-references GB:L36132; NID:g556287; PID:g777773
REFERENCE I48608
#authors Thuvesson, M.; Albrecht, D.; Zurcher, G.; Andres, A.C.;
#journal Biochem. Biophys. Res. Commun. (1995) 209:582-589
#title iyk, a novel intracellular protein tyrosine kinase
#title differentially expressed in the mouse mammary gland and
#title intestine.
#cross-references MUID:95251656
#accession I48608
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-153, 'T', 155-236, 'H', 238-512 ##label RB2
#cross-references EMBL:248757; NID:g736263; PID:g736264
GENETICS
#gene BSK
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
#keywords homology; SH2 homology; SH3 homology
#feature ATP; intestine; phosphotransferase
FEATURE
56-112 #domain SH3 homology #label SH3\

```

```

123-215 #domain SH2 homology #label SH2\
239-501 #domain protein kinase homology #label KIN\
247-255 #region protein kinase ATP-binding motif
SUMMARY #length 512 #molecular-weight 58891 #checks 3691
Query Match 39.5%; Score 276; DB 2; Length 512;
Best Local Similarity 37.6%; Pred. No. 1.29e-33;
Matches 35; Conservative 26; Mismatches 32; Indels 0; Gaps 0;
Db 123 WFGAIKRAADAKQLLYSENQTGAFLIRSESSOKGDFSLSDVDEGVVKKHYRIRRLDEGGF 182
QY 120 WYFGSVSRTOAQLLSPNPEGAFILRPSESSIGGYSLSVRAQAKVCHYRVSMAADGSL 179
Db 183 FLTRRKVFSTLNEFVNYTTTSDGLCVKLEKPC 215
QY 180 YLQKGRLPFGLELLTYKANKWLQNPLQPC 212
RESULT 4
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) 4 - freshwater sponge
ALTERNATE_NAMES (Spongilla lacustris)
ORGANISM src-type tyrosine kinase 4
#formal_name Spongilla lacustris
DATE 07-May-1993 #sequence_revision 07-May-1993 #text_change
08-Sep-1997
ACCESSIONS S24553
REFERENCE S24550
#authors Raulf, F.
#submission submitted to the EMBL Data Library, September 1991
#accession S24553
#molecule_type mRNA
#residues 1-506 ##label RAU
#cross-references EMBL:X61604; NID:g10155; PID:g10156
GENETICS
#gene srk4
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
#keywords homology; SH2 homology; SH3 homology
#feature ATP; phosphotransferase; tyrosine-specific protein kinase
FEATURE
61-111 #domain SH3 homology #label SH3\
122-214 #domain SH2 homology #label SH2\
238-496 #domain protein kinase homology #label KIN\
246-254 #region protein kinase ATP-binding motif\
268 #active_site Lys #status predicted
SUMMARY #length 506 #molecular-weight 57561 #checks 9002
Query Match 38.5%; Score 269; DB 2; Length 506;
Best Local Similarity 40.5%; Pred. No. 3.01e-32;
Matches 32; Conservative 24; Mismatches 21; Indels 2; Gaps 2;
Db 122 WFGQVKRVDAEKQLMMPNNLGSFLIRSDTTPGDFSLSDVDRVHYRIKLENGTY 181
QY 120 WYFGSVSRTOAQLLSPNPEGAFILRPSESSIGGYSLSVRAQAKVCHYRVSMAADGSL 179
Db 182 FVTR-RLTFQSIQELVAY 199
QY 180 YLQKGRLPFGLELLTY 197
RESULT 5
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) 1 - freshwater sponge
ALTERNATE_NAMES (Spongilla lacustris)
ORGANISM src-type tyrosine kinase 1
#formal_name Spongilla lacustris
DATE 07-May-1993 #sequence_revision 07-May-1993 #text_change
08-Sep-1997
ACCESSIONS S24550
REFERENCE S24550
#authors Raulf, F.
#submission submitted to the EMBL Data Library, September 1991
#accession S24550

```

```

##molecule_type mRNA
##residues 1-505 ##label RAU
##cross-references EMBL:X61601; NID:g10149; PID:g10150
GENETICS
#gene srk1
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
KEYWORDS ATP; phosphotransferase; SH3 homology
FEATURE
61-111 #domain SH3 homology #label SH3\
122-214 #domain SH2 homology #label SH2\
238-496 #domain protein kinase homology #label KIN\
246-254 #region protein kinase ATP-binding motif\
268 #active_site Lys #status predicted
SUMMARY #length 505 #molecular-weight 57693 #checksum 3389

Query Match 37.5%; Score 262; DB 2; Length 505;
Best Local Similarity 39.8%; Pred. No. 6.94e-31;
Matches 37; Conservative 21; Mismatches 35; Indels 0; Gaps 0;

Db 122 WFLGKIKRVEAEKMLNQSFNQGSLIRDSETPGDFSLSKVDQDRVHRHYRVRLEDGSL 181
QY 120 WYFGSVSRTOAQLLSPNPGAFILRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179
Db 182 FVTRRSFQIHLHLDVHYKIETDGLCCCKLLYPC 214
QY 180 YLQGRLPFGLEELTYKANKWKLQNPLQPC 212

RESULT 6
ENTRY I38396 #type complete
TITLE Protein-tyrosine kinase (EC 2.7.1.112) FRK - human
ALTERNATE_NAMES FYN-related kinase (FRK)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change
09-Apr-1998

ACCESSIONS I38396
REFERENCE Lee, J.; Wang, Z.; Luoh, S.M.; Wood, W.I.; Scadden, D.T.
#journal Gene (1994) 138:247-251
#title Cloning of FRK, a novel intracellular SRC-like tyrosine
kinase-encoding gene.
#cross-references MUID:94171047
#accession I38396
#status preliminary
#molecule_type mRNA
##residues 1-505 ##label RES
##cross-references EMBL:U00803; NID:g392887; PID:g392888
GENETICS
#gene GDB:FRK
#map_position 4q35-4q35
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
KEYWORDS ATP; phosphotransferase
FEATURE
49-105 #domain SH3 homology #label SH3\
232-494 #domain protein kinase homology #label KIN\
240-248 #region protein kinase ATP-binding motif
SUMMARY #length 505 #molecular-weight 58254 #checksum 9379

Query Match 37.2%; Score 260; DB 2; Length 505;
Best Local Similarity 41.0%; Pred. No. 1.70e-30;
Matches 32; Conservative 23; Mismatches 23; Indels 0; Gaps 0;

Db 116 WFGAIGRDAEKOLLVSENTGTGFLIRESEKQGFSLVDGAVVHYRIKRLDEGGF 175
QY 120 WYFGSVSRTOAQLLSPNPGAFILRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179
Db 176 FLTRRRFSTLNEFVSHY 193
QY 180 YLQGRLPFGLEELTY 197

```

---

```

RESULT 7
ENTRY OKHULK #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) lck - human
ALTERNATE_NAMES kinase-related transforming protein (lck)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
22-May-1998

ACCESSIONS J00152; S07822; S07200; S01879; S07143; A32797; I57636
REFERENCE J00152
#authors Rouer, E.; Van Huynh, T.; de Souza, S.L.; Lang, M.C.;
Fischer, S.; Benarous, R.
#journal Gene (1989) 84:105-113
#title Structure of the human lck gene: differences in genomic
organisation within src-related genes affect only
N-terminal exons.
#cross-references MUID:90108697
#accession J00152
##molecule_type DNA
##residues 1-509 ##label ROU
##cross-references EMBL:X14053
REFERENCE S07822
#authors Perlmutter, R.M.; Marth, J.D.; Lewis, D.B.; Peet, R.;
Ziegler, S.F.; Wilson, C.B.
#journal J. Cell. Biochem. (1988) 38:117-126
#title Structure and expression of lck transcripts in human lymphoid
cells.
#cross-references MUID:89123626
#accession S07822
##molecule_type mRNA
##residues 1-86, 'P', 88-509 ##label PER
##cross-references EMBL:X13529; NID:g34294; PID:g34295
REFERENCE S07200
#authors Koga, Y.; Caccia, N.; Toyonaga, B.; Spolski, R.; Yanagi, Y.;
Yoshikai, Y.; Mak, T.W.
#journal Eur. J. Immunol. (1986) 16:1643-1646
#title A human T cell-specific cDNA clone (YTI6) encodes a protein
with extensive homology to a family of protein-tyrosine
kinases.
#cross-references MUID:87133831
#accession S07200
##molecule_type mRNA
##residues 1-205, 'ASAITPI', 212-257, 'RCGW', 262, 'TTT', 266, 'T',
268-281, 'AGRLP', 287-503, 'STA' ##label KOG
REFERENCE S01879
#authors Veillette, A.; Foss, F.M.; Sausville, E.A.; Bolen, J.B.;
Rosen, N.
#journal Oncogene Res. (1987) 1:357-374
#title Expression of the lck tyrosine kinase gene in human colon
carcinoma and other non-lymphoid human tumor cell lines.
#cross-references MUID:88217332
#accession S01879
##molecule_type mRNA
##residues 368-471, 'H', 473-509 ##label VEI
##cross-references EMBL:X06369; NID:g34288; PID:g34289
REFERENCE S07143
#authors Trevillyan, J.M.; Lin, Y.; Chen, S.J.; Phillips, C.A.; Canna,
C.; Linna, T.J.
#journal Biochim. Biophys. Acta (1986) 888:286-295
#title Human T lymphocytes express a protein-tyrosine kinase
homologous to p56(LSTRA).
#cross-references MUID:87000726
#accession S07143
##molecule_type mRNA
##residues 'A', 376-509 ##label TRE
##cross-references EMBL:X04476; NID:g35779; PID:g35780
REFERENCE A32797
#authors Takadera, T.; Leung, S.; Gernone, A.; Koga, Y.; Takiyama, Y.;
Miyamoto, N.G.; Mak, T.W.
#journal Mol. Cell. Biol. (1989) 9:2173-2180
#title Structure of the two promoters of the human lck gene:
differential accumulation of two classes of lck transcripts

```

[illegible]

Db 148 WYFGKITRRESERLLNPENPRGTFLVRESEETKGAYCLSVDFDNAKGLNVKHYKIRKL 207

**KEYWORDS**

homology; SH2 homology; SH3 homology  
alternative splicing; ATP; autophosphorylation; blocked amino end;  
lipoprotein; myristylation; phosphopeptide; phosphorylase; protein kinase  
phosphotransferase; proto-oncogene; tyrosine-specific protein kinase

---

**FEATURE**

88-137 148-245 265-523 273-281 2  12, 48 17 34, 46 72 295 527  <b>SUMMARY</b>	#domain SH3 homology #label SH3\ #domain SH2 homology #label SH2\ #domain protein kinase homology #label KIN\ #region protein kinase ATP-binding motif\ #modified_site myristylated amino end (Gly) (in mature form) #status predicted\ #binding_site phosphate (Ser) (covalent) (by protein kinase C) #status predicted\ #binding_site phosphate (Ser) (covalent) (by protein kinase A) #status predicted\ #binding_site phosphate (Thr) (covalent) #status experimental\ #binding_site phosphate (Ser) (covalent) #status experimental\ #active_site Lys #status predicted\ #binding_site phosphate (Tyr) (covalent) #status predicted #length 533 #molecular_weight 60023 #checksum 238
--	--

Query Match            33.6%; Score 235; DB 1; Length 533;  
Best Local Similarity 38.6%; Pred. No. 1.10e-25;  
Matches         32; Conservative      26; Mismatches 20; Indels    5; Gaps    3;

---

Dbb    148 WYFGKTRRSERLLLNPENPGTFVRETTKGAYCLSVDFDNAGLNKYHKIRKL 207  
Qy    120 WYFSGVRTQAQLLPPNPEGAFILRPESLSGGYSLSVR-A-Q-AK---VCYRVSMNA 174  
  
Ddb    208 DSGGFYTSTRQTFSLIQQLVAVY 230  
Qy    175 ADGLYLQKGRLFPELETTY 197  
|::: |::::: ||  
|::: |::::: ||

---

```

RESULT   13
ENTRY    TVFVS2           #type complete
TITLE    protein-tyrosine kinase (EC 2.7.1.112) src - avian sarcoma virus S2
ORGANISM #formal_name avian sarcoma virus S2
DATE     31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 07-Nov-1997
ACCESSION B25375
REFERENCE A25375
AUTHORS   Ikawa, S.; Hagino-Yamagishi, K.; Kawai, S.; Yamamoto, T.; Toyoshima, K.
JOURNAL   Mol. Cell. Biol. (1986) 6:2420-2428
TITLE     Activation of the cellular src gene by transducing retrovirus.
```

---

```
#cross-references MUId:B7064539
#accession       B25375
#molecule_type DNA
#residues        1-557 #label IKA

GENETICS
#gene            src
CLASSIFICATION   #superfamily protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH3 homology
KEYWORDS         ATP; autophosphorylation; blocked amino end; lipoprotein; myristylation; oncogene; phosphoprotein; phosphotransferase; transforming protein; tyrosine-specific protein kinase
```

---

**FEATURE**

88-137 148-245 265-523 273-281 2  295 416	#domain SH3 homology #label SH3\ #domain SH2 homology #label SH2\ #domain protein kinase homology #label KIN\ #region protein kinase ATP-binding motif\ #modified_site myristylated amino end (Gly) (in mature form) #status predicted\ #binding_site phosphate (Tyr) (covalent) (by active-site Lys #status predicted) #binding_site phosphate (Tyr) (covalent) (by binding-site Lys #status predicted)
--	---

	SUMMARY	#length 557 #molecular-weight 62582 #checksum 7810 Query Match 33.6%; Score 235; DB 1; Length 557; Best Local Similarity 38.6%; Pred. NO. 1.10e-25; Matches 32; Conservative 26; Mismatches 20; Indels 5; Gaps 3;
Db	148 WYFGKITRRESERLLNPNRGTFVRESETHKAYCLSVSDFDNAGLNVKHKKRL 207  :::  :::       I ::    : II  :::	
Qy	120 WYFSGVRTAQQLSPNPGEAFILRPSSSLGGYSLSVRA-Q-AK---VCHRYVSMA 174 I:::  :::       I ::    : II  :::	
Db	208 DSCGFYTTSRTQSFLSQVLWAY 230 I:::  :::       I ::    : II  :::	
Qy	175 ADGSLYLKGRLFPGLLETTY 197 I:::  :::       I ::    : II  :::	
RESULT	14 TVEVPD #type complete ENTRY protein-tyrosine kinase.(EC 2.7.1.112) src _avian_sarcoma TITLE virus PR2257 ORGANISM #formal_name avian sarcoma virus PR2257 DATE 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 07-Nov-1997	
ACCESSIONS	A30174	
REFERENCE	A30174 Jeryk, J.; Dezelee, P.; Barnier, J.V.; Svoboda, J.; Rehyba, J.; Karakoz, I.; Rynditch, A.V.; Yatsula, B.A.; Calothy, G. #journal J. Virol. (1989) 63:481-492 #title Transduction of the cellular src gene and 3' adjacent sequences in avian sarcoma virus PR2257.	
#cross-references	MUID:8904972	
#accession	A30174	
#molecule_type	DNA	
#residues	1-587 #label GER	
GENETICS	src #gene superfamily protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH3 homology ATP; autophosphorylation; blocked amino end; lipoprotein; myristylation; oncogene; phosphoprotein; phosphotransferase; transforming protein; tyrosine-specific protein kinase	
CLASSIFICATION	#domain SH3 homology #label SH3\ #domain SH2 homology #label SH2\ #domain protein kinase homology #label KIN\ #region protein kinase ATP-binding motif\ #modified_site myristylated amino end (Gly) (in mature form) #status predicted\ #active-site Lys-#status: predicted\ #binding_site phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted	
KEYWORDS	nucleic acid sequence not shown; not compared with conceptual translation	
FEATURE	88-137 #domain SH3 homology #label SH3\ 148-245 #domain SH2 homology #label SH2\ 265-523 #domain protein kinase homology #label KIN\ 273-281 #region protein kinase ATP-binding motif\ 2 #modified_site myristylated amino end (Gly) (in mature form) #status predicted 295 #active-site Lys-#status: predicted 416 #binding_site phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted	
SUMMARY	#length 587 #molecular-weight 65800 #checksum 2040 Query Match 33.6%; Score 235; DB 1; Length 587; Best Local Similarity 38.6%; Pred. No. 1.10e-25; Matches 32; Conservative 26; Mismatches 20; Indels 5; Gaps 3;	
Db	148 WYFGKITRRESERLLNPNRGTFVRESETHKAYCLSVSDFDNAGLNVKHKKRL 207  :::  :::       I ::    : II  :::	
Qy	120 WYFSGVRTAQQLSPNPGEAFILRPSSSLGGYSLSVRA-Q-AK---VCHRYVSMA 174 I:::  :::       I ::    : II  :::	
Db	208 DSCGFYTTSRTQSFLSQVLWAY 230 I:::  :::       I ::    : II  :::	
Qy	175 ADGSLYLKGRLFPGLLETTY 197 I:::  :::       I ::    : II  :::	
RESULT	15 TVHUHC #type complete ENTRY protein-tyrosine kinase (EC 2.7.1.112) hck - human TITLE ORGANISM #formal_name Homo sapiens #common_name man DATE 31-Dec-1989 #sequence_revision 10-Nov-1995 #text_change 22-May-1998	



```

269      predicted\
390      #active_site Lys #status predicted\
      #binding_site phosphate (Tyr) (covalent) (by
SUMMARY      #length 505 #molecular-weight 57312 #checksum 7314
      #length 505 #molecular-weight 57312 #checksum 7314

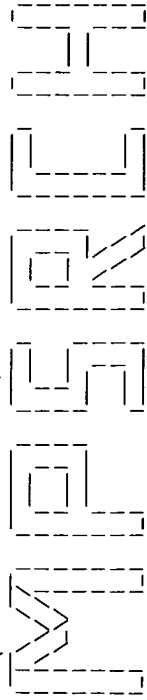
Query Match      33.3%; Score 233; DB 1; Length 505;
Best Local Similarity 41.4%; Pred. No. 2.64e-25;
Matches 41; Conservative 24; Mismatches 27; Indels 7; Gaps 5;

Db 123 WFEKGISRKDAEROLLAPGNMIGSFMRDSEITKGSYLSVRDYPDROGDTVKHKYKIRTL 182
Qy 120 WIFSGVSRTOAQQLLSPNPEGAFIRPSESLSGYSLSVR---A-QA-KVCHYRVSWA 174

Db 183 DNGGFYISPRSTFTIQELVDHYKKGNDGLCOK-LSVPC 220
Qy 175 ADGSLYLQKGRLEPGLLELLTYK-ANWKLIONPLLOPC 212
```

Search completed: Thu May 20 12:55:42 1999  
Job time : 122 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:56:02 1999; MasPar time 5.35 Seconds  
Tabular output not generated. 491.220 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (120-212) from US09099053.pap (4 of 6)  
Perfect Score: 699  
Sequence: 1 WYFGVSRVTAQQLLLSPN.....LTYKANKWKLQNPQLQPC 93

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 40.582; Variance 67.112; scale 0.605

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	625	89.4	496	1 SRM_MOUSE	TYROSINE-PROTEIN KINAS	2.01e-122
2	269	38.5	505	1 SRK1_SPOLA	TYROSINE-PROTEIN KINAS	1.64e-37
3	262	37.5	505	1 SRK1_SPOLA	TYROSINE-PROTEIN KINAS	6.09e-36
4	260	37.2	505	1 FRK_HUMAN	TYROSINE-PROTEIN KINAS	1.71e-35
5	241	34.5	508	1 LCK_MOUSE	PROTO-ONCOGENE TYROSIN	2.83e-31
6	238	34.0	508	1 LCK_MOUSE	PROTO-ONCOGENE TYROSIN	1.30e-30
7	236	33.8	552	1 SRC1_DROME	TYROSINE-PROTEIN KINAS	3.57e-30
8	235	33.6	526	1 SRC_RSVSR	TYROSINE-PROTEIN KINAS	5.92e-30
9	235	33.6	526	1 SRC_AVISR	TYROSINE-PROTEIN KINAS	5.92e-30
10	235	33.6	532	1 SRC_CHICK	PROTO-ONCOGENE TYROSIN	5.92e-30
11	235	33.6	557	1 SRC_AVIS2	TYROSINE-PROTEIN KINAS	5.92e-30
12	235	33.6	587	1 SRC_AVIS2	TYROSINE-PROTEIN KINAS	5.92e-30
13	233	33.3	526	1 HCK_HUMAN	TYROSINE-PROTEIN KINAS	1.62e-29
14	233	33.3	544	1 YES_XIPHE	PROTO-ONCOGENE TYROSIN	1.62e-29
15	232	33.2	509	1 SRK_HYDAT	TYROSINE-PROTEIN KINAS	2.69e-29
16	232	33.2	568	1 SRC_AVIS	TYROSINE-PROTEIN KINAS	2.69e-29
17	230	32.9	498	1 BLK_MOUSE	TYROSINE-PROTEIN KINAS	7.36e-29
18	230	32.9	503	1 HCK_RAT	TYROSINE-PROTEIN KINAS	7.36e-29
19	229	32.8	503	1 SRC_RSVPA	TYROSINE-PROTEIN KINAS	1.22e-28
20	227	32.5	523	1 HCK_MOUSE	TYROSINE-PROTEIN KINAS	3.33e-28
21	227	32.5	541	1 YES_CHICK	PROTO-ONCOGENE TYROSIN	3.33e-28
22	226	32.3	526	1 SRC_PSVH1	TYROSINE-PROTEIN KINAS	5.49e-28
23	225	32.2	535	1 SRC_HUMAN	PROTO-ONCOGENE TYROSIN	9.06e-28

24	225	32.2	537	1 YES_XENLA	PROTO-ONCOGENE TYROSIN	9.06e-28
25	225	32.2	540	1 SRCN_MOUSE	NEURONAL PROTO-ONCOGEN	9.06e-28
26	224	32.0	529	1 FGR_HUMAN	PROTO-ONCOGENE TYROSIN	1.50e-27
27	223	31.9	507	1 LCK_CHICK	PROTO-ONCOGENE TYROSIN	2.47e-27
28	223	31.9	526	1 SRC_RSVR	TYROSINE-PROTEIN KINAS	2.47e-27
29	222	31.8	228	1 SEM5_CABEL	SEX MUSCLE ABNORMAL PR	4.07e-27
30	221	31.6	541	1 YES_MOUSE	PROTO-ONCOGENE TYROSIN	6.70e-27
31	220	31.5	543	1 YES_HUMAN	PROTO-ONCOGENE TYROSIN	1.10e-26
32	219	31.3	746	1 ABL_MLVAB	TYROSINE-PROTEIN KINAS	1.82e-26
33	219	31.3	1123	1 ABL_MOUSE	PROTO-ONCOGENE TYROSIN	1.82e-26
34	219	31.3	1130	1 ABL1_HUMAN	PROTO-ONCOGENE TYROSIN	1.82e-26
35	218	31.2	211	1 DRK_DROME	PROTEIN E(SEV)2B (SH2-	2.99e-26
36	217	31.0	217	1 GRB2_HUMAN	GROWTH FACTOR RECEPTOR	4.91e-26
37	217	31.0	217	1 GRB2_MOUSE	GROWTH FACTOR RECEPTOR	4.91e-26
38	217	31.0	439	1 ABL_FSVHY	TYROSINE-PROTEIN KINAS	4.91e-26
39	217	31.0	1182	1 ABL2_HUMAN	TYROSINE-PROTEIN KINAS	4.91e-26
40	216	30.9	545	1 FGR_FSVGR	TYROSINE-PROTEIN KINAS	8.08e-26
41	215	30.8	531	1 SRC2_XENLA	TYROSINE-PROTEIN KINAS	1.33e-25
42	215	30.8	531	1 SRC1_XENLA	TYROSINE-PROTEIN KINAS	1.33e-25
43	213	30.5	217	1 GRB2_XENLA	GROWTH FACTOR RECEPTOR	3.58e-25
44	213	30.5	1252	1 PIP5_HUMAN	1-PHOSPHATIDYLINOSITOL	3.58e-25
45	212	30.3	517	1 FGR_MOUSE	PROTO-ONCOGENE TYROSIN	5.87e-25

ALIGNMENTS

RESULT	ID	SRM_MOUSE	STANDARD;	PRT;	496 AA.
AC	062270;	062270;			
DT	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	TYROSINE-PROTEIN KINASE SRM (EC 2.7.1.112) (PTK70).				
GN	SRMS OR SRM.				
OS	MUS MUSCULUS (MOUSE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-C57BL/6; TISSUE-THYMUS;				
RX	MEDLINE; 97369678.				
RX	KAWACHI Y., NAKAUCHI H., OTSUKA F.;				
RT	"Isolation of a cDNA encoding a tyrosine kinase expressed in murine				
RT	skin."				
RL	EXP. DERMATOL. 21:533-538(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-LUNG;				
RX	MEDLINE; 95021220.				
RA	KOHMURA N., YAGI T., TOMOOKA Y., OYANAGI M., KOMINAMI R., TAKEDA N.,				
RA	CHIBA J., IKAWA Y., AIZAWA S.;				
RT	"A novel nonreceptor tyrosine kinase, Srm: cloning and targeted				
RT	disruption."				
CC	MOL. CELL. BIOL. 14:6915-6925(1994).				
CC	-!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +				
CC	PROTEIN TYROSINE PHOSPHATE.				
CC	-!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC				
CC	DOMAIN. BELONGS TO THE SRC SUBFAMILY.				
CC	-!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.				
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; D49427; G684972; -				
DR	EMBL; D26186; G529073; -				
DR	MGI; MGI:101865; SRMS.				
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.				

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS0001; SH2; 1.  
 DR PROSITE; PS0002; SH3; 1.  
 DR PFAM; PF00017; SH2; 1.  
 DR PFAM; PF00018; SH3; 1.  
 DR PFAM; PF00069; pkinase; 1.  
 DR HSP; P11362; IFGL.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
 KW SH3 DOMAIN; PHOSPHORYLATION.  
 FT DOMAIN 55 116 SH3.  
 FT DOMAIN 124 216 SH2.  
 FT DOMAIN 234 495 PROTEIN KINASE.  
 FT NP\_BIND 240 248 ATP (BY SIMILARITY).  
 FT BINDING 262 262 ATP (BY SIMILARITY).  
 FT ACT\_SITE 354 354 BY SIMILARITY.  
 FT MOD\_RES 384 384 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CONFLICT 78 78 R -> G (IN REF. 2).  
 FT CONFLICT 236 238 LRK -> FGR (IN REF. 2).  
 FT CONFLICT 278 278 N -> I (IN REF. 2).  
 SQ SEQUENCE 496 AA; 55731 MW; FD44DEF6 CRC32;

Query Match 89.4%; Score 625; DB 1; Length 496;  
 Best Local Similarity 82.8%; Pred. No. 2.01e-122;  
 Matches 77; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

Db 124 WFGSGISRAQAQQLLSANAPAGFLIRSESSIGGYSLSVRAQAKVCHYRICMAPSGSL 183  
 QY 120 WFGSGVSRQAQQLLSPPNEPGAFLIRSESSIGGYSLSVRAQAKVCHYRVSMADGSL 179

Db 184 YLOEGQLFSLDALLAYKTNKLIQNPLQPC 216  
 QY 180 YLQKGRLEPGLLELLYYRANWKLIQNPLQPC 212

RESULT 2  
 ID SRK4\_SPOLA STANDARD; PRT; 506 AA.  
 AC P42690;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TYROSINE-PROTEIN KINASE SRK4 (EC 2.7.1.112).  
 GN SRK4.  
 OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).  
 OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;  
 OC HAPLOSLERIDA; SPONGILLIDAE; SPONGILLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92334872.  
 RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;  
 RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge  
 Spongilla lacustris.";  
 RT ONCOGENE 7:1625-1630(1992).  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -!- ALTERNATIVE PRODUCTS: SRK1 AND SRK4 MAY ARISE BY ALTERNATIVE  
 CC SPLICING.  
 CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X61604; G10156; -  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS0001; SH2; 1.  
 DR PROSITE; PS0002; SH3; 1.

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS0001; SH2; 1.  
 DR PROSITE; PS0002; SH3; 1.  
 DR PFAM; PF00017; SH2; 1.  
 DR PFAM; PF00018; SH3; 1.  
 DR PFAM; PF00069; pkinase; 1.  
 DR HSP; P00523; 2PTK.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
 KW SH3 DOMAIN; PHOSPHORYLATION.  
 FT DOMAIN 54 116 SH3.  
 FT DOMAIN 122 214 SH2.  
 FT NP\_BIND 240 254 ATP (BY SIMILARITY).  
 FT BINDING 268 268 ATP (BY SIMILARITY).  
 FT ACT\_SITE 359 359 BY SIMILARITY.  
 SQ SEQUENCE 506 AA; 57561 MW; 807A71D0 CRC32;

Query Match 38.5%; Score 269; DB 1; Length 506;  
 Best Local Similarity 40.5%; Pred. No. 1.64e-37;  
 Matches 32; Conservative 24; Mismatches 21; Indels 2; Gaps 2;

Db 122 WFTGQVRYVDAEKQLMPPFNLSFLIRSDTTPGDFSLSVRDIDRYRHYRIKLENGTY 181  
 QY 120 WFTSGVSRQAQQLLSPPNEPGAFLIRSESSIGGYSLSVRAQAKVCHYRVSMADGSL 179

Db 182 FVTR-RLTFOSIOELVAYY 199  
 QY 180 YLQKGRLEPGLLELLYY 197

RESULT 3  
 ID SRK1\_SPOLA STANDARD; PRT; 505 AA.  
 AC P42686;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TYROSINE-PROTEIN KINASE SRK1 (EC 2.7.1.112).  
 GN SRK1.  
 OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).  
 OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;  
 OC HAPLOSLERIDA; SPONGILLIDAE; SPONGILLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92334872.  
 RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;  
 RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge  
 Spongilla lacustris.";  
 RT ONCOGENE 7:1625-1630(1992).  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -!- ALTERNATIVE PRODUCTS: SRK1 AND SRK4 MAY ARISE BY ALTERNATIVE  
 CC SPLICING.  
 CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X61601; G10150; -  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS0001; SH2; 1.  
 DR PROSITE; PS0002; SH3; 1.

DR PFAM: PF00017; SH2: 1.  
 DR PFAM: PF00018; SH3: 1.  
 DR PFAM: PF00069; pkinase; 1.  
 DR HSP: P00523; 2PTK.  
 KW SH3 DOMAIN; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
 FT DOMAIN 54 116 SH3.  
 FT DOMAIN 122 214 SH2.  
 FT DOMAIN 240 493 PROTEIN KINASE.  
 FT NP\_BIND 246 234 ATP (BY SIMILARITY).  
 FT BINDING 268 268 ATP (BY SIMILARITY).  
 FT ACT\_SITE 359 359 BY SIMILARITY.  
 SQ SEQUENCE 505 AA; 57693 MW; 3AE3DF34 CRC32;

Query Match 37.58; Score 262; DB 1; Length 505;  
 Best Local Similarity 39.84; Pred. No. 6.09e-36;  
 Matches 37; Conservative 21; Mismatches 35; Indels 0; Gaps 0;

Db 122 WELGKIKRVEAEKMLQSNQVSEFLRDSSTPGDFSLVKDQDRVHYRVRLEDGSL 181  
 QY 120 WIFSGVSRTOAQQLLSPPNEPGAFLIRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179  
 Db 182 FVTRRTFTQILHVDHYKIETDGLCKLLYPC 214  
 QY 180 YLQKGRFLPGLELLTYYYKANWLIQNPLQPC 212

RESULT 4  
 ID FRK\_HUMAN STANDARD; PRT; 505 AA.  
 AC P42685; Q13128;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE TYROSINE-PROTEIN KINASE FRK (EC 2.7.1.112) (NUCLEAR TYROSINE PROTEIN KINASE RAK).  
 GN FRK.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LYMPHOID;  
 RX MEDLINE; 94171047.  
 RA LEE J., WANG Z., LUOH S.-M., WOOD W.I., SCADDEN D.T.;  
 RT "Cloning of FRK, a novel human intracellular SRC-like tyrosine kinase-encoding gene";  
 RL GENE 138:247-251(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95210168.  
 RA CANCE W.G., CRAVEN R.J., BERGMAN M., XU L.H., ALITALO K., LIU E.T.;  
 RT "Rak, a novel nuclear tyrosine kinase expressed in epithelial cells.";  
 RL CELL GROWTH DIFFER. 5:1347-1355(1994).  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE; 93293373.  
 RA CANCE W.G., CRAVEN R.J., WEINER T.M., LIU E.T.;  
 RT "Novel protein kinases expressed in human breast cancer.";  
 RL INT. J. CANCER 54:571-577(1993).  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: RESTRICTED TO CELLS LINES DERIVED FROM TISSUES  
 CC OF LYMPHOID, BRAIN, BLADDER COLON AND BLADDER ORIGIN.  
 CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL: U00803; G392888; -.  
 DR EMBL: U22322; G732528; -.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00011; PROTEIN KINASE\_DOM; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00069; pkinase; 1.  
 DR HSP: P00523; 2PTK.  
 KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
 FT DOMAIN 42 110 SH3.  
 FT DOMAIN 116 208 SH2.  
 FT NP\_BIND 234 491 PROTEIN KINASE.  
 FT BINDING 262 262 ATP (BY SIMILARITY).  
 FT ACT\_SITE 354 354 ATP (BY SIMILARITY).  
 FT MOD\_RES 387 387 BY SIMILARITY.  
 FT VARIANT 122 122 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CONFLICT 115 115 G -> R.  
 SQ SEQUENCE 505 AA; 58254 MW; C4226A83 CRC32;

Query Match 37.24; Score 260; DB 1; Length 505;  
 Best Local Similarity 41.04; Pred. No. 1.71e-35;  
 Matches 32; Conservative 23; Mismatches 23; Indels 0; Gaps 0;

Db 116 WFFGAIGRSDAEKOLYSENKGTSGFLIRESEKQSFSLVDGAVVKKRIKRLDEGGF 175  
 QY 120 WIFSGVSRTOAQQLLSPPNEPGAFLIRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179  
 Db 176 FLTRRTFTSLNEFVSHY 193  
 QY 180 YLQKGRFLPGLELLTY 197

RESULT 5  
 ID LCK\_HUMAN STANDARD; PRT; 508 AA.  
 AC P06239; P07100;  
 DT 01-JAN-1988 (REL. 06, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE LCK (EC 2.7.1.112) (P56-LCK)  
 DE (LSK) (T CELL-SPECIFIC PROTEIN-TYROSINE KINASE).  
 GN LCK.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89123626.  
 RA PERLMUTTER R.M., MARTH J.D., LEWIS D.B., PEET R., ZIEGLER S.F.,  
 RA WILSON C.B.;  
 RT "Structure and expression of lck transcripts in human lymphoid  
 RT cells.";  
 RL J. CELL. BIOCHEM. 38:117-126(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87133831.  
 RA KOGA Y., CACCIA N., TOYONAGA B., SPOLSKI R., YANAGI Y., YOSHIKAI Y.,  
 RA MAK T.W.;  
 RT "A human T cell-specific cDNA clone (Yt16) encodes a protein with  
 RT extensive homology to a family of protein-tyrosine kinases";  
 RL EUR. J. IMMUNOL. 16:1643-1646(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90108697.  
 RA ROUER E., VAN HUYNH T., DE SOUZA S.L., LANG M.C., FISCHER S.,



RESULT 6  
 ID LCK\_MOUSE STANDARD; PRT; 508 AA.  
 AC P06240; Q61794; Q61795; Q62320;  
 DT 01-JAN-1988 (REL. 06, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE LCK (EC 2.7.1.112) (P56-LCK)  
 GN LCK OR LSK-T.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 86079521.  
 RA MARTIN J.D., PEET R., KREBS E.G., PERLMUTTER R.M.;  
 RT "A lymphocyte-specific protein-tyrosine kinase gene is rearranged and  
 RT overexpressed in the murine T cell lymphoma LSTRA.";  
 RL CELL 43:393-404(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 86146842.  
 RA VORONOVA A.F., SEFTON B.M.;  
 RT "Expression of a new tyrosine protein kinase is stimulated by  
 RT retrovirus promoter insertion.";  
 RL NATURE 319:682-685(1986).  
 RN [3]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 89096891.  
 RA GARVIN A.M., PAWAR S., MARTIN J.D., PERLMUTTER R.M.;  
 RT "Structure of the murine lck gene and its rearrangement in a murine  
 RT lymphoma cell line.";  
 RL MOL. CELL. BIOL. 8:3058-3064(1988).  
 RN [4]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RX MEDLINE; 88142832.  
 RA VORONOVA A.F., ADLER H.T., SEFTON B.M.;  
 RT "Two lck transcripts containing different 5' untranslated regions are  
 RT present in T cells.";  
 RL MOL. CELL. BIOL. 7:4407-4413(1987).  
 RN [5]  
 RP MUTAGENESIS OF TYR-504.  
 RX MEDLINE; 88248001.  
 RA AMREIN K.E., SEFTON B.M.;  
 RT "Avian reovirus mRNAs are nonfunctional in infected mouse cells:  
 RT translational basis for virus host-range restriction.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 85:4257-4261(1988).  
 RN [6]  
 RP MUTAGENESIS.  
 RX MEDLINE; 93059694.  
 RA HURLEY T.R., AMREIN K.E., SEFTON B.M.;  
 RT "Creation and characterization of temperature-sensitive mutants of  
 RT the lck tyrosine protein kinase.";  
 RL J. VIROL. 66:7406-7413(1992).  
 RN [7]  
 RP MUTAGENESIS OF LYS-272.  
 RX MEDLINE; 91163633.  
 RA ABRAHAM N., MICELI M.C., PARNES J.C., VEILLETTA A.;  
 RT "Enhancement of T-cell responsiveness by the lymphocyte-specific  
 RT tyrosine protein kinase p56lck.";  
 RL NATURE 350:62-66(1991).  
 RN [8]  
 RP MUTAGENESIS OF TYR-504.  
 RX MEDLINE; 91219495.  
 RA ABRAHAM K.M., LEVIN S.D., MARTIN J.D., FORBUSH K.A., PERLMUTTER R.M.;  
 RT "Thymic tumorigenesis induced by overexpression of p56lck.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:3977-3981(1991).  
 RN [9]  
 RP MUTAGENESIS.  
 RX MEDLINE; 93133805.  
 RA CARRERA A.C., ALEXANDROV K., ROBERTS T.M.;  
 RT "The conserved lysine of the catalytic domain of protein kinases is  
 RT actively involved in the phosphotransfer reaction and not required  
 RT for anchoring ATP.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:442-446(1993).  
 RN [10]  
 RP PALMITOYLATION.  
 RX MEDLINE; 94019312.  
 RA SHENOY-SCARIA A.M., TIMSON L.K., KWONG J., SHAW A.S., LUBLIN D.M.;  
 RT "Palmitoylation of an amino-terminal cysteine motif of protein  
 RT tyrosine kinases p56lck and p59fyn mediates interaction with  
 RT glycosyl-phosphatidylinositol-anchored proteins.";  
 RL MOL. CELL. BIOL. 13:6385-6392(1993).  
 RN [11]  
 RP PALMITOYLATION.  
 RX MEDLINE; 95071286.  
 RA KOEGL M., ZLATKINE P., LEY S.C., COURTNEIDGE S.A., MAGEE A.I.;  
 RT "Palmitoylation of multiple Src-family kinases at a homologous N-  
 RT terminal motif.";  
 RL BIOCHEM. J. 303:749-753(1994).  
 CC -!- FUNCTION: MAY PARTICIPATE IN ANTIGEN-INDUCED T-CELL ACTIVATION.  
 CC ITS EARLY EXPRESSION IS ESSENTIAL FOR EARLY T-LYMPHOCYTE  
 CC DEVELOPMENT.  
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -!- ENZYME REGULATION: REGULATED BY PHOSPHORYLATION ON TYR-504.  
 CC -!- SUBCELLULAR LOCATION: BOUND TO THE CYTOPLASMIC DOMAIN OF EITHER  
 CC CD4 OR CD8.  
 CC -!- TISSUE SPECIFICITY: PRESENT AT A LOW LEVEL IN MOST T CELLS, AND  
 CC AT AN ELEVATED LEVEL IN LSTRA AND THY 19 (T-CELL LYMPHOMA) CELLS.  
 CC -!- DEVELOPMENTAL STAGE: LEVELS REMAIN RELATIVELY CONSTANT THROUGHOUT  
 CC T-CELL ONTOGENY.  
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X03533; G54814; -  
 DR EMBL; M12056; G198764; -  
 DR EMBL; X03533; E7690; ALT\_SEQ.  
 DR EMBL; X03533; E7691; ALT\_SEQ.  
 DR EMBL; M21511; G554186; ALT\_SEQ.  
 DR EMBL; M18098; G198767; -  
 DR PIR; A23639; A23639.  
 DR MGD; MGI:96756; LCK.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 DR PFAM; PF00017; SH2; 1.  
 DR PFAM; PF00018; SH3; 1.  
 DR PFAM; PF00069; pkinase; 1.  
 DR HSP; P06239; 1LCK.  
 KW PROTO-ONCOGENE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; TRANSFERASE;  
 KW ATP-BINDING; MYRISTYLATION; SH2 DOMAIN; SH3 DOMAIN; PALMITATE;  
 KW LIPOPROTEIN.  
 FT INIT\_MET 0 0 PROBABLE.  
 FT LIPID 1 1 MYRISTATE (BY SIMILARITY).  
 FT LIPID 2 2 PALMITATE.  
 FT LIPID 4 4 PALMITATE.  
 FT DOMAIN 60 120 SH3.  
 FT DOMAIN 126 223 SH2.  
 FT DOMAIN 244 497 PROTEIN KINASE.  
 FT NP\_BIND 250 258 ATP (BY SIMILARITY).  
 FT BINDING 272 272 ATP (BY SIMILARITY).

```

FT ACT_SITE 363 363 BY SIMILARITY.
FT MOD_RES 393 393 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 504 504 PHOSPHORYLATION (POTENTIAL).
FT MUTAGEN 268 268 K->N: REDUCED ACTIVITY.
FT MUTAGEN 269 269 V->L: REDUCED ACTIVITY.
FT MUTAGEN 270 270 A->S: REDUCED ACTIVITY.
FT MUTAGEN 271 271 V->A: REDUCED ACTIVITY.
FT MUTAGEN 272 272 S->R: LOSS OF ACTIVITY.
FT MUTAGEN 273 273 S->N: REDUCED ACTIVITY.
FT MUTAGEN 274 274 L->M: REDUCED ACTIVITY.
FT MUTAGEN 275 275 K->V: REDUCED ACTIVITY.
FT MUTAGEN 504 504 Y->F: CAUSES THYMIC TUMORS.
FT CONFLICT 282 283 VP-> DA (IN REF. 2).
SQ SEQUENCE 508 AA; 57821 MW; 70423E50 CRC32;

Query Match 34.0%; Score 238; DB 1; Length 508;
Best Local Similarity 39.4%; Pred. No. 1.30e-30;
Matches 39; Conservative 22; Mismatches 31; Indels 7; Gaps 4;

Db 126 WFFKNSLRDAERQLLAPGNTGHSFLIRESESTAGSFLSVDFOQNGEVVYKIRNL 185
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Qy 120 WYFSGVSRTOAQQLLSPNPEGAFLIRPSESSLGYSLSVRA---QAKVC-HYRVSM 174
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

Db 186 DNGGYI-SPRTFFGLDLVHYNASDGLCTKLSRPC-223
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Qy 175 ADGSLYLQKGR-LFPGLELLTYRANKWKLQNPLQLQC 212
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

RESULT 7
ID SRC_LDROME STANDARD; PRT; 552 AA.
AC P00528;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRC64B (EC 2.7.1.112).
GN SRC1.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RX MEDLINE; 86028179.
RA SIMON M.A., DREES B., KORNBERG T., BISHOP J.M.;
RT "The nucleotide sequence and the tissue-specific expression of
RT Drosophila c-src.";
RL CELL 42:831-840(1985).
RN [2]
RX SEQUENCE OF 249-552 FROM N.A.
RX MEDLINE; 84082064.
RA HOFFMANN F.M., PRESCO L.D., HOFFMAN-FALK H., SHILO B.-Z.;
RT "Nucleotide sequences of the Drosophila src and abl homologs:
RT conservation and variability in the src family oncogenes.";
RL CELL 35:393-401(1983).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11917; G158501; -
CC EMBL; K01043; G157304; -
CC PIR; A00634; TVFFS.
CC FLYBASE; FBgn0003501; Src64B.

```

```

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P11362; IFGI.
DR TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; PHOSPHORYLATION;
KW TRANSFERASE; ATP-BINDING; SH3 DOMAIN; SH2 DOMAIN.
FT DOMAIN 95 156 SH3.
FT DOMAIN 162 259 SH2.
FT DOMAIN 284 537 PROTEIN KINASE.
FT NP_BIND 290 298 ATP (BY SIMILARITY).
FT BINDING 312 312 ATP (BY SIMILARITY).
FT ACT_SITE 404 404 BY SIMILARITY.
FT MOD_RES 434 434 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 261 263 KPQ -> ASL (IN REF. 2).
FT CONFLICT 266 269 MWDL -> TAAPDVGF (IN REF. 2).
FT CONFLICT 272 272 E -> Q (IN REF. 2).
FT CONFLICT 286 287 LL -> VV (IN REF. 2).
FT CONFLICT 290 290 L -> V (IN REF. 2).
FT CONFLICT 293 293 G -> R (IN REF. 2).
FT CONFLICT 316 316 E -> A (IN REF. 2).
FT CONFLICT 366 366 D -> N (IN REF. 2).
FT CONFLICT 373 373 G -> D (IN REF. 2).
FT CONFLICT 384 385 IA -> MH (IN REF. 2).
FT CONFLICT 389 390 AS -> TT (IN REF. 2).
FT CONFLICT 393 393 E -> Q (IN REF. 2).
FT CONFLICT 400 400 L -> V (IN REF. 2).
FT CONFLICT 406 407 AA -> TT (IN REF. 2).
FT CONFLICT 435 435 C -> R (IN REF. 2).
FT CONFLICT 471 471 M -> T (IN REF. 2).
FT CONFLICT 484 484 M -> L (IN REF. 2).
FT CONFLICT 507 507 F -> L (IN REF. 2).
FT CONFLICT 536 536 F -> L (IN REF. 2).
SQ SEQUENCE 552 AA; 63018 MW; C2A73FD0 CRC32;

Query Match 33.8%; Score 236; DB 1; Length 552;
Best Local Similarity 38.4%; Pred. No. 3.57e-30;
Matches 38; Conservative 22; Mismatches 32; Indels 7; Gaps 5;

Db 162 WFFENVLRKADKLLAEENRGTFLVRPSEHNPNNGYSLSVKDWDGCGYHVKYRIKPL 221
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Qy 120 WYFSGVSRTOAQQLLSPNPEGAFLIRPSESSLGYSLSVRA-Q-AK---VCHYRVSM 174
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

Db 222 DNGGYIATNQTFFSLQALVMAYSKNALGLCHI-LSRPC 259
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Qy 175 ADGSLYLQKGR-LFPGLEEL-LTYRANKWKLQNPLQLQC 212
|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

RESULT 8
ID SRC_RSISR STANDARD; PRT; 526 AA.
AC P00524;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TRANSFORMING PROTEIN SRC (BC 2.7.1.112) (P60-
DE SRC).
DE V-SRC.
OS ROUS SARCOMA VIRUS (STRAIN SCHMIDT-RUPPIN).
OC VIRUSES; RETROVIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 89160256.
RA BARNIER J.V., DEZELBE P., MARX M., CALOTHY G.;
RT "Nucleotide sequence of the src gene of the Schmidt-Ruppin strain of
RT ROUS sarcoma virus type E.";
RL NUCLEIC ACIDS RES. 17:1252-1252(1989).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE; 83141780.

```

[illegible]









[illegible]

```
RESULT 15
ID STK_HYDAT STANDARD; PRT: 509 AA.
AC P17713;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE STK (EC 2.7.1.112) (P57-STK).
GN STK.
OS HYDRA ATTENUATA (HYDRA) (HYDRA VULGARIS).
OC EUKARYOTA; METAZOA; CNIDARIA; HYDROZOA; HYDROIDA; ANTHOMEDUSAE;
OC HYDRIDAE; HYDRA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90066418.
RA BOSCH T.C.G., UNGER T.F., FISHER D.A., STEELE R.E.;
RT "Structure and expression of STK, a src-related gene in the simple
RT metazoan Hydra attenuata.";
RL MOL. CELL. BIOL. 9:4141-4151(1989).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25245; G159274; -
DR PIR; A34094; TVHAST.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P00523; 1PRL.
KW TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; PHOSPHORYLATION;
KW TRANSFERASE; ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
FT DOMAIN 59 120 SH3.
FT DOMAIN 126 218 SH2.
FT DOMAIN 240 495 PROTEIN KINASE.
FT NP_BIND 246 254 ATP (BY SIMILARITY).
FT BINDING 288 288 ATP (BY SIMILARITY).
FT ACT_SITE 360 360 BY SIMILARITY.
FT MOD_RES 390 390 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 509 AA; 56885 MW; 2B724CE9 CRC32;
Query Match 33.2%; Score 232; DB 1; Length 509;
Best Local Similarity 37.2%; Pred. No. 2.69e-29;
Matches 35; Conservative 20; Mismatches 37; Indels 2; Gaps 2;
Db 126 WYFGDVKRAEAKRLMVRGLPGSTFLIRKAETAAGVNFSLSVRDGDSVKHYVRKLDTG 185
QY 120 WYFSGVSRTOAQQLLSPNPGAFILRSESSLGYSLSVRAQAKVCHYRVSNAADGSL 179
Db 186 FITTRAPFNSLYELVQHYTKDADGLV-CALTLP 218
QY 180 YLQKGRLEPFGLELLTY-KANWKLIQNPLQPC 212
```

Search completed: Thu May 20 12:56:17 1999  
Job time : 15 secs.

\*\*\*\*\*

WIREH

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, G.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 12:56:35 1999; Maspar time 13.40 seconds  
378.732 Million cell updates/sec

Tabular output not generated.

Title: >US-09-099-053-2  
Description: (120-212) from US09099053.pep (4 of 6)  
Perfect Score: 699

Sequence: 1 WYFSGVSRQAQQLLSPPN.....LLTYRKANWKLQNPLQPC 93

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb19

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 39.060; Variance 68.598; scale 0.569

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	276	39.5	512 11 Q61364	B-CELL SRC-HOMOLOGY TY	4.85e-37
2	274	39.2	512 11 Q61745	B-CELL SRC-HOMOLOGY TY	1.30e-36
3	272	38.9	506 11 Q62662	SRC RELATED TYROSINE K	3.48e-36
4	264	37.8	517 5 Q94879	DSRC41.	1.76e-34
5	260	37.2	496 13 Q93411	NON-RECEPTOR PROTEIN T	1.25e-33
6	252	36.1	451 11 Q64434	SRC-RELATED INTESINAL	6.14e-32
7	241	34.5	363 4 Q13152	P56LCK.	1.26e-29
8	241	34.5	512 4 Q12850	LYMPHOCYTE-SPECIFIC PR	1.26e-29
9	240	34.3	451 4 Q13882	TYROSINE KINASE.	2.04e-29
10	235	33.6	523 14 Q85477	SRC.	2.25e-28
11	235	33.6	526 14 Q93080	TSUP1 SRC (FRAGMENT).	2.25e-28
12	235	33.6	526 14 Q64994	(SCHMIDT-RUPPIN D STRA	2.25e-28
13	235	33.6	533 13 Q90992	C-SRC.	2.25e-28
14	235	33.6	533 13 Q98915	GENE C-SRC PRODUCING P	2.25e-28
15	235	33.6	535 14 Q92957	SRC TYROSINE KINASE.	2.25e-28
16	234	33.5	212 4 Q43726	DJ370M22.1 (GROWTH FAC	3.64e-28
17	234	33.5	330 4 Q73791	GADS PROTEIN.	3.64e-28
18	230	32.9	587 14 Q64817	PROTEIN-TYROSINE KINAS	2.47e-27
19	229	32.8	322 11 Q89100	GRB-2-RELATED MONOCYTI	3.98e-27
20	229	32.8	545 14 Q86362	PP62V.	3.98e-27

21	229	32.8	546 14 Q86363	PP62V.	3.98e-27
22	226	32.3	526 11 Q60567	H-19 PROVIRAL SEQUENCE	1.67e-26
23	225	32.2	526 14 Q64993	TYROSINE KINASE.	2.68e-26
24	223	31.9	488 13 Q13064	LYN PROTEIN TYROSINE K	6.94e-26
25	223	31.9	525 14 Q92806	P60 SRC.	6.94e-26
26	220	31.5	281 11 Q60898	SRC-LIKE ADAPTER PROTE	2.88e-25
27	219	31.3	980 14 Q92809	P120 POLYPROTEIN.	4.62e-25
28	219	31.3	981 14 Q39477	P120 GAG-ABL POLYPROTE	4.62e-25
29	219	31.3	1149 4 Q13869	PROTO-ONCOGENE TYROSIN	4.62e-25
30	216	30.9	276 4 Q13239	PUTATIVE SRC-LIKE ADAP	1.91e-24
31	216	30.9	663 14 Q39851	GARDNER-RASHEED FELINE	1.91e-24
32	212	30.2	517 11 Q61404	GARDNER-RASHEED FELINE	1.26e-23
33	211	30.2	377 13 P79956	SH2/SH3 ADAPTOR PROTEI	2.01e-23
34	211	30.2	352 6 Q28414	GAG-ONC FUSION PROTEIN	2.01e-23
35	211	30.2	812 14 Q85466	COMPLETE GENOME.	2.01e-23
36	210	30.0	505 4 Q16291	BLK=PROTEIN TYROSINE K	3.22e-23
37	207	29.6	526 14 Q07461	TYROSINE-PROTEIN KINAS	1.31e-22
38	205	29.3	410 5 Q24218	SH2/SH3 ADAPTOR PROTEI	3.34e-22
39	205	29.3	450 13 Q73786	C-SRC KINASE.	3.34e-22
40	204	29.2	517 11 Q63206	FRK MRNA.	5.33e-22
41	199	28.5	377 11 Q55032	SH2/SH3 ADAPTOR PROTEI	5.45e-21
42	198	28.3	534 4 Q16248	P59FYF.	8.66e-21
43	198	28.3	537 11 Q62844	PROTO-ONCOGENE FYN.	8.66e-21
44	194	27.8	615 11 Q63614	TYROSINE KINASE RECEPT	5.50e-20
45	193	27.6	597 11 Q64509	PROTEIN TYROSINE PHOSP	8.72e-20

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	512 AA.
ID	Q61364			
AC	Q61364;			
DT	01-NOV-1996 (TREMBREL. 01, CREATED)			
DT	01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)			
DE	B-CELL SRC-HOMOLOGY TYROSINE KINASE (PROTEIN TYROSINE KINASE).			
GN	FRK OR BSK.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 95137395.			
RA	OBORG-WELSH C.; WELSH M.;			
RT	"Cloning of BSK, a murine FRK homologue with a specific pattern of			
RT	tissue distribution";			
RL	GENE 152:239-242(1995).			
DR	EMBL; L36132; G777773; -.			
DR	MGI; MGI:103265; FRK.			
DR	PROSITE; PS00107; PROTEIN_KINASE_TYR; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PFAM; PF00017; SH2; 1.			
DR	PFAM; PF00018; SH3; 1.			
DR	PFAM; PF00069; pkinase; 1.			
SQ	SEQUENCE 512 AA; 5891 MW; DECFS3C7 CRC32;			
Query Match 39.5%; Score 276; DB 11; Length 512;				
Best Local Similarity 37.8%; Pred. No. 4.85e-37;				
Matches 35; Conservative 26; Mismatches 32; Indels 0; Gaps 0;				
DB	123 WFFGAIKRAAEKQLLYSENQTCAGFLIRESEKQDFSLSVLDELGVVVKHYRIRRLDEGGF 182			
QY	120 WYFSGVSRQAQQLLSPPNPEAGFLIRPSESSLGYSLSVRAQAKVCHIRVSMADGSL 179			
DB	183 FTTRRVFTSLNEFVNYTTTSDGLCVKLEKPC 215			
QY	180 YLQKGLRFLPGLELLTYRKANWKLQNPLQPC 212			
RESULT	2			
ID	Q61745			
AC	Q61745;			
DT	01-NOV-1996 (TREMBREL. 01, CREATED)			

```
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE B-CELL SRC-HOMOLOGY TYROSINE KINASE (INTESTINAL TYROSINE KINASE).
GN FRK OR IYK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC CUCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE: 95251656
RA THUVESON M., ALBRECHT D., ZUERCHER G., ANDRES A., ZIEMIECKI A.;
RT "Iyk, a novel intracellular protein tyrosine kinase differentially
expressed in the mouse mammary gland and intestine.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 209:582-589(1995).
DR EMBL: 248757; G736264; -.
DR MGD: MGI:103265; FRK.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
SQ SEQUENCE 512 AA; 58928 MW; 808D1612 CRC32;

Query Match 39.2%; Score 274; DB 11; Length 512;
Best Local Similarity 36.6%; Pred. No. 1.30e-36;
Matches 34; Conservative 27; Mismatches 32; Indels 0; Gaps 0;

Db 123 WFGAIAKRAEAKQQLLYSENOTGAFLIRSEKOGDLSVLDEGVVVKHYRIRLDEGGF 182
QY 120 WYFGSVSRTOAQQLLSPNPPGAFILRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179
Db 183 FLTRRRVFTSLNEFVNYITTSDDGLCVKLEKPC 215
QY 180 YLQKGRLPGLLELLYYKANWKLQNPLLOPC 212

RESULT 3
ID Q62662 PRELIMINARY; PRT; 506 AA.
AC Q62662;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SRC RELATED TYROSINE KINASE.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC CUCUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE OF 356-416 FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RA SUNITHA I., AVIGAN M.I.;
RL BIOCHIM. BIOPHYS. ACTA, GENE STRUCT. EXPR. 1221:348-352(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RX MEDLINE: 96330334.
RA SUNITHA I., AVIGAN M.I.;
RT "The apical membranes of maturing gut columnar epithelial cells
contain the enzymatically active form of a newly identified
tyr-related tyrosine kinase.";
RL ONCOGENE 13:547-559(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RA AVIGAN M.I.;
RL SUBMITTED (MAY-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U09583; G939625; -.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
SQ SEQUENCE 506 AA; 58166 MW; 8A3172C1 CRC32;
```

```
Query Match 38.9%; Score 272; DB 11; Length 506;
Best Local Similarity 37.6%; Pred. No. 3.48e-36;
Matches 35; Conservative 25; Mismatches 33; Indels 0; Gaps 0;

Db 117 WFGAIAKRAEAKQQLLYSENOTGAFLIRSEKOGDLSVLDEGVVVKHYRIRLDEGGF 176
QY 120 WYFGSVSRTOAQQLLSPNPPGAFILRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179
Db 177 FLTRRRVFTSLNEFVNYITTSDDGLCVKLEKPC 209
QY 180 YLQKGRLPGLLELLYYKANWKLQNPLLOPC 212

RESULT 4
ID Q94879 PRELIMINARY; PRT; 517 AA.
AC Q94879;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE DSR41.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON S.
RX MEDLINE: 96268448.
RA TAKAHASHI F., ENDO S., KOJIMA T., SAIGO K.;
RT "Regulation of cell-cell contacts in developing Drosophila eyes by
Dsr41, a new, close relative of vertebrate c-src.";
RL GENES DEV. 10:1645-1656(1996).
DR EMBL: D42125; D1008290; -.
DR FLYBASE: FBgn0004603; Src41.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
SQ SEQUENCE 517 AA; 59097 MW; A33DC5EA CRC32;

Query Match 37.8%; Score 264; DB 5; Length 517;
Best Local Similarity 44.9%; Pred. No. 1.76e-34;
Matches 35; Conservative 12; Mismatches 31; Indels 0; Gaps 0;

Db 130 WYPRKIKRIEAKKLLLPENEGHGAFLIRDSERHNDYSLSVRGDVTYKHVRIROLDEGGF 189
QY 120 WYFGSVSRTOAQQLLSPNPPGAFILRPSESLGGYSLSVRAQAKVCHYRVSMADGSL 179
Db 190 FIARTTFTLQELVEHY 207
QY 180 YLQKGRLPGLLELLTY 197

RESULT 5
ID Q93411 PRELIMINARY; PRT; 496 AA.
AC Q93411;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NON-RECEPTOR PROTEIN TYROSINE KINASE LALOO.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA WEINSTEIN D.C., MARDEN J., CARNEVALI F., HEMMATI-BRIVANLOU A.;
RT "FGF-mediated mesoderm induction involves the Src-family kinase
laloo.";
RL NATURE 0:0-0(1998).
DR EMBL: AF081803; G3411274; -.
SQ SEQUENCE 496 AA; 56275 MW; BC2C7DA3 CRC32;

Query Match 37.2%; Score 260; DB 13; Length 496;
```

ID	Q13152	PRELIMINARY;	PRT;	363 AA.
AC	Q13152;			
DT	01-NOV-1996	(TREMBLREL, 01, CREATED)		
DT	01-NOV-1996	(TREMBLREL, 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL, 08, LAST ANNOTATION UPDATE)		
DE	P561CK.			
GN	LCK.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
EN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE; 94067101.			
RA	VOGEL L.B., FUJITA D.J.;			
RT	"The SH3 domain of p56lck is involved in binding to			
RT	phosphatidylinositol 3'-kinase from T lymphocytes.";			
RL	MOL. CELL. BIOL. 13:7408-7417(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE; 95155308.			
RA	VOGEL L.B., FUJITA D.J.;			
RT	"p70 phosphorylation and binding to p56lck is an early event in			
RT	interleukin-2-induced onset of cell cycle progression in			
RT	T-lymphocytes.";			
RL	J. BIOL. CHEM. 270:2506-2511(1995).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE; 96085119.			
RA	VOGEL L.B., ARTHUR R., FUJITA D.F.;			
RT	"An aberrant lck mRNA in two human T-cell lines.";			
RL	BIOTCHIM. BIOPHYS. ACTA 1264:168-172(1995).			
DR	EMBL; U23852; G775208; -			
DR	PFAM; PF00017; SH2; 1.			
DR	PFAM; PF00018; SH3; 1.			
DR	PFAM; PF00069; pkinase; 1.			
SQ	SEQUENCE 363 AA; 40866 MW; D2822D6A CRC32;			
	Query Match 34.5%; Score 241; DB 4; Length 363;			
	Best Local Similarity 40.4%; Pred. No. 1.26e-29;			
	Matches 40; Conservative 21; Mismatches 31; Indels 7; Gaps 4			
Db	127 WFFNLSRKDAERQLAPGNTHGSLFIRESESTAGSFLSRVDFDQNGGEVVKHYKIRNL 186			
Qy	120 WYFGSVSRKTAQQLLSPNPENPAGFLRPSSESLGGYSLSVRA---QAKVC-HYRVYMA 174			
Db	187 DNGEYI-SPRITPGLHVLVRYHTNASDGLCTRLSRPC 224			
Qy	175 ADGSLYLQKGR-L-EPGLELLTYKANKWKLQNPLLPQC 212			
RESULT	8			
ID	Q12850	PRELIMINARY;	PRT;	512 AA.
AC	Q12850;			
DT	01-NOV-1996	(TREMBLREL, 01, CREATED)		
DT	01-NOV-1996	(TREMBLREL, 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL, 08, LAST ANNOTATION UPDATE)		
DE	LYMPHOCYTE-SPECIFIC PROTEIN TYROSINE KINASE.			
GN	LCK.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
EN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LEUKEMIA;			
FX	MEDLINE; 94187714.			
RA	WRIGHT D.D., SEFTON B.M., KAMPS M.P.;			
RT	"Oncogenic activation of the Lck protein accompanies translocation of			
RT	the Lck gene in the human HSB2 T-cell leukemia.";			
RL	MOL. CELL. BIOL. 14:2429-2437(1994).			
DR	EMBL; U07236; G460966; -			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
PFAM	PF00017; SH2.1			



```

DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 512 AA; 58412 MW; 12BOBA65 CRC32;

Query Match      34.5%; Score 241; DB 4; Length 512;
Best Local Similarity 40.4%; Pred.No. 1.26e-29;
Matches 40; Conservative 21; Mismatches 31; Indels 7; Gaps 4;

Db 127 WFTKLSRKDAERQLLAPGNTHGSLFIRESSTAGSFLSVSRFDQNGEVVVKYKIRNL 186
   1 : | | : | | : | | : | | : | | | | | : | | | | | : | | : | | :
Qy 120 WYFGVSRTQAQQLLSPNPPEGAFLIRPSESSILGGYSLSVRA---QAKVC-HYRVSM A 174
   1 : | | : | | : | | : | | | | | : | | | | | : | | : | | :

Db 187 DNGCFYI-SPIRIFPGHLHVRHYTNASDGLCTRLSRPC 224
   : | | : | | : | | | | | : | | : | | :
Qy 175 ADGSLYLQKGL-FPGLEELTYTKANKWKLIONPLLPQC 212
   : | | : | | : | | | | | : | | : | | :

RESULT          9
ID Q13882; PRELIMINARY; PRT; 451 AA.
AC Q13882;
DT DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DI TYROSINE KINASE.
GN BRK OR PK6.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
[1]
RN RN SEQUENCE FROM N.A.
RP RP TISSUE-BREAST CARCINOMA;
RC MEDLINE; 94309916.
RX MITCHELL P.J., BARKER K.T., MARTINDALE J.E.;
RT "Cloning and characterisation of cDNAs encoding a novel non-receptor
RT tyrosine kinase, brk, expressed in human breast tumours.";
RL ONCOGENE 9:2383-2390(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97430836.
RY PARK S.H., LEE K.-H., KIM H., LEE S.-T.;
RT "Assignment of the human PK6 gene encoding a non-receptor protein
RT tyrosine kinase to 20q13.3 by fluorescence in situ hybridization.";
RL CYTOGENET. CELL GENET. 77:271-272(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98419955.
RY LEE H., KIM M., LEE K.-H., KANG K.-N., LEE S.-T.;
RT "Exon-intron structure of the human PK6 gene demonstrates that PK6
RT constitutes a distinct family of non-receptor tyrosine kinase.";
RL MOL. CELLS 8:401-407(1998).
DR DR ENBL; X78549; G515026; -.
DR DR ENBL; U61412; G3551753; -.
DR DR ENBL; U61406; G3551753; JOINED.
DR DR ENBL; U61407; G3551753; JOINED.
DR DR ENBL; U61408; G3551753; JOINED.
DR DR ENBL; U61409; G3551753; JOINED.
DR DR ENBL; U61410; G3551753; JOINED.
DR DR ENBL; U61411; G3551753; JOINED.
DR DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 451 AA; 51834 MW; 4AC30408 CRC32;

Query Match      34.3%; Score 240; DB 4; Length 451;
Best Local Similarity 39.4%; Pred.No. 2.04e-29;
Matches 37; Conservative 16; Mismatches 39; Indels 2; Gaps 2;

Db 78 WFTGCISRSBAVRRLQAEQNATGAFLIRVEKPSADYLVSRDQTAVRYHKIWRAGRL 137
   1 : | | : | | : | | : | | | | | : | | | | | : | | : | | :
Qy 120 WYFGVSRTQAQQLLSPNPPEGAFLIRPSESSILGGYSLSVRAQAKVCHYRVSMADGSL 179
   1 : | | : | | : | | : | | | | | : | | | | | : | | : | | :

```



```

RESULT 15
ID O92957 PRELIMINARY; PRT; 535 AA.
AC O92957;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SRC TYROSINE KINASE.
GN SRC.
OS ROUS SARCOMA VIRUS.
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCHMIDT-RUPPIN B;
RA BOUCK J., SKALKA A.M., KATZ R.A.;
RT "Complete nucleotide sequence of avian sarcoma virus.";
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AF052428; G3003004; -.
SQ SEQUENCE 535 AA; 60042 MW; 78365B79 CRC32;

Query Match 33.6%; Score 235; DB 14; Length 535;
Best Local Similarity 38.6%; Pred. No. 2,25e-28;
Matches 32; Conservative 26; Mismatches 20; Indels 5; Gaps 3;

Db 148 WYFGKITRRESERLLNPNPRGTFVRESEETKAYCLSVSDFDNAGLNKYKIRKL 207
Qy 120 WYFGSVSRTOAQQLLLSPNPGALIRPSESSLGYSLSVRA-Q-AK---VCHYRVSWA 174

Db 208 DSGFYITSRTQFSSLOQLVAYY 230
Qy 175 ADGSLYLQKGRFLPGLELLTY 197

```

Search completed: Thu May 20 12:59:38 1999  
 Job time : 183 secs.

\*\*\*\*\*

WORLD

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 13:00:28 1999; Maspar time 20.61 seconds  
Tabular output not generated. .... 259:028 Million cell-updates/sec

Title: >US-09-099-053-2  
Description: (230-480) from US09099053.pap (5 of 6)  
Perfect Score: 1882  
Sequence: 1 FALGRKLGEGYFGEVWGLW.....ECWRSSPERPSPATLREKL 251  
Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 33.669; Variance 147.292; scale 0.229

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the\_total\_score\_distribution...

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1882	100.0	488	39	W89248 Human SAD.	4.02e-168
2	1033	54.9	451	12	R63088 Breast tumor kinase,	1.41e-85
3	1019	54.1	505	8	R41941 PTK gene LPTK-2 prod.	3.15e-84
4	1019	54.1	505	14	R85929 Protein tyrosine-kinase	3.15e-84
5	997	53.0	533	8	R39705 Chicken pp60 c-src pr	4.16e-82
6	993	52.8	536	8	R39706 Human pp60 c-src prot	1.01e-81
7	962	51.1	259	6	R32299 Sequence of pp60(c-sr	9.78e-79
8	935	49.7	417	3	R42021 (Beta-galactosidase N	3.87e-76
9	831	44.2	1146	3	R15156 Abelson Related Gene,	3.73e-66
10	799	42.5	441	17	R94538 Drosophila Src28C tyr	4.31e-63
11	785	41.7	1182	3	R15157 Abelson Related gene,	9.40e-62
12	769	40.9	466	13	R71132 N-terminal truncated	3.18e-60
13	769	40.9	507	15	R84181 Megakaryocyte kinase	3.18e-60
14	769	40.9	507	13	R71133 Cytoplasmic tyrosine k	3.18e-60
15	769	40.9	630	17	R94536 TEC tyrosine kinase.	3.18e-60
16	765	40.6	620	17	R94535 ITK tyrosine kinase.	7.66e-60

17	757	40.2	246	13	R71131 Tyrosine kinase domai	4.44e-59
18	749	39.8	675	17	R94533 BMX tyrosine kinase.	2.58e-58
19	749	39.8	675	15	R84182 Megakaryocyte kinase	2.58e-58
20	747	39.7	659	17	R94534 BTK tyrosine kinase.	4.00e-58
21	744	39.5	659	20	W06708 Mouse haematopoietic-	7.73e-58
22	742	39.4	928	19	R97853 Rat REK7 eph-related	1.20e-57
23	738	39.2	877	35	W71628 Mouse Bsk receptor-l1	2.89e-57
24	735	39.1	991	15	R85090 Eph-like receptor pro	5.58e-57
25	735	39.1	1005	38	W83147 Rat receptor tyrosine	5.58e-57
26	730	38.8	710	13	R75714 Eph-related PTK Cdk7,	1.67e-56
27	730	38.8	722	13	R75705 Eph-related PTK Cdk7,	1.67e-56
28	730	38.8	744	13	R75713 HEK polypeptide.	1.67e-56
29	724	38.5	983	6	R31466 Human mack protein.	6.24e-56
30	721	38.3	528	34	W64454 Elk PTK.	1.20e-55
31	718	38.2	380	8	R44512 Mouse developmenta	2.33e-55
32	718	38.2	984	8	R44513 Mouse developmenta	2.33e-55
33	715	38.0	998	15	W03421 Eph-like receptor pr	4.49e-55
34	714	37.9	998	15	R85092 Eph-related Cdk5.	5.59e-55
35	708	37.6	951	13	R75704 Eph-related Cdk5.	2.08e-54
36	708	37.6	994	24	W26366 Mouse Nuk tyrosine k	2.08e-54
37	707	37.6	995	13	R75712 Eph-related PTK Cdk5.	2.59e-54
38	707	37.6	1011	13	R75709 Eph-related PTK Cdk5+	2.59e-54
39	705	37.5	994	15	R87018 Receptor tyrosine kin	4.02e-54
40	703	37.4	849	13	R75706 Eph-related PTK Cdk8.	6.23e-54
41	704	37.4	983	13	R75711 Eph-related PTK Cdk4.	5.01e-54
42	703	37.4	986	14	R85936 Protein tyrosine-kin	6.23e-54
43	703	37.4	986	15	R85091 Eph-like receptor pro	6.23e-54
44	702	37.3	970	15	R85089 Eph-like receptor pro	7.76e-54
45	693	36.8	977	23	W19258 Embryonic stem cell k	5.57e-53

ALIGNMENTS

RESULT 1

ID	W89248	standard; Protein; 488 AA.
AC	W89248:	
DE	Human SAD.	
DE	10-MAR-1999	(first entry)
KW	PTP04;	PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW	type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;	
KW	neurodegenerative disease; neuronal survival; Alzheimer's disease;	
KW	Parkinson's disease; Huntington's disease.	
OS	Homo sapiens.	
PN	W09849317-A2.	
PD	05-NOV-1998.	
PF	27-APR-1998;	U08439.
PR	23-OCT-1997;	US-063595.
PR	28-APR-1997;	US-044428.
PR	20-MAY-1997;	US-047222.
PR	11-JUN-1997;	US-049477.
PR	11-JUN-1997;	US-049756.
PR	18-JUN-1997;	US-049914.
PA	(SUGEN) SUGEN INC.	
PI	App H, Clary D, Courtneidge SA, Hui TH, Jallal B,	
PI	Markby D, Onrust S, Peles E, Plowman GD;	
DR	WPI: 99-009434/01.	
DR	N-PSDB: V81743.	
PT	New nucleic acid encoding specific protein tyrosine phosphatases -	
PT	useful for identifying specific modulators for treatment and	
PT	prevention of cancer and neurodegenerative disease	
PS	Claim 2: Page 154-155; 193pp; English.	
CC	The present invention describes isolated, enriched or purified nucleic	
CC	acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The	
CC	present sequence represents human SAD. The above proteins, other than	
CC	ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify	
CC	substances that modulate their activity (i.e. agonists and antagonists,	
CC	including NBP) in vivo or in vitro. These substances are used to treat	
CC	or prevent diseases associated with abnormal signal transduction	
CC	pathways that involve the proteins, particularly cancer (e.g. leukaemia	
CC	and lymphoma), while modulators of ALK-7 (which is a type I receptor	
CC	serine/threonine kinase) are used to promote neuronal survival,	
CC	particularly for treating Alzheimer's, Parkinson's or Huntington's	
CC	diseases. Nucleic acid fragments of the polynucleotides encoding the	



QY 292 CSQGEPIVITELMRKGNLQALFGLTPEGRALRLPPLLLGFACQVAEGMSYLEEQRVVHRDL 351  
 Db 356 aarnvlvgehnlykvdafglarvkfdnediyeshrheiklpvkwtapeairsnkfsiksd 415  
 QY 352 AARNVLVDGLACKVADFGGLARLK-D--DIYSPSSSSKIPVKWTAPFAAANYRVFSOKSD 408  
 Db 416 vwsfgillyeilitygmypsgmtgagvqlmqlagnyrlpqpsncpqgqfynlmlecnwaepek 475  
 QY 409 VWSFGVLLHEVFTYGQCPYEGMTNHETLQOIMRGYLRPFAACPAEVYVLMLECWRSPE 468  
 Db 476 erptfetlrwkl 487  
 QY 469 ERPSFATLREKL 480

## RESULT 4

ID R85929 standard; Protein; 505 AA.  
 AC R85929;  
 DT 14-FEB-1996 (first entry)  
 DE Protein tyrosine-kinase LptK2.  
 KW Protein tyrosine-kinase; pTK; LptK2; agonist; cell growth;  
 OS Homo sapiens.  
 PN WO9527061-A1.  
 PD 12-OCT-1995.  
 PF 04-APR-1995; U04228.  
 PR 04-APR-1994; US-222616.  
 PA (GETH ) GENENTECH INC.  
 PI Bennett BD, Goeddel D, Lee JM, Matthews W, Tsai SP;  
 PI Wood WI.  
 DR WPI; 95-366160/47.  
 DR N-PSDB; T03097.  
 PT Agonist antibodies which activate specific protein tyrosine  
 PT kinase(s) - also activate chimeric proteins of kinase extracellular  
 PT domain and Ig constant domain, useful for studying, and therapeutic  
 PT modulation of, cell growth and differentiation  
 PS Disclosure: Page 56-58; 125pp; English.  
 CC DNA probes based on protein tyrosine-kinase (pTK) sequences were used  
 CC to screen cDNA libraries to identify novel pTK genes. A LptK2 gene  
 CC (T03097) was isolated from lymphocytic and megakaryocytic cell  
 CC libraries. The encoded novel pTK, LptK2 (R85929), shows homology to  
 CC known pTKs, and can be used to design drugs that modulate pTK  
 CC activity.  
 SQ Sequence 505 AA;

Query Match 54.1%; Score 1019; DB 14; Length 505;  
 Best Local Similarity 53.6%; Pred. No. 3.15e-84;  
 Matches 135; Conservative 43; Mismatches 71; Indels 3; Gaps 2;

Db 236 llkrlggfgevewglwnnttpvavtklpgsgmdpndfiraegimknlrpkltgiyav 295  
 QY 232 LGRKLGEGYGEVWEGWGLWGLSLPVAIKVKSANKMLTDLAKEIOTLKGLRHERLIRHAV 291  
 Db 296 ctledpiyitelmrhgslqeylqndtgskihltqqvdmagvasgmaylestrnyihrdl 355  
 QY 292 CSQGEPIVITELMRKGNLQALFGLTPEGRALRLPPLLLGFACQVAEGMSYLEEQRVVHRDL 351  
 Db 356 aarnvlvgehnlykvdafglarvkfdnediyeshrheiklpvkwtapeairsnkfsiksd 415  
 QY 352 AARNVLVDGLACKVADFGGLARLK-D--DIYSPSSSSKIPVKWTAPFAAANYRVFSOKSD 408  
 Db 416 vwsfgillyeilitygmypsgmtgagvqlmqlagnyrlpqpsncpqgqfynlmlecnwaepek 475  
 QY 409 VWSFGVLLHEVFTYGQCPYEGMTNHETLQOIMRGYLRPFAACPAEVYVLMLECWRSPE 468  
 Db 476 erptfetlrwkl 487  
 QY 469 ERPSFATLREKL 480

## RESULT 5

ID R39705 standard; Protein; 533 AA.  
 AC R39705;

DT 23-DEC-1993 (first entry)  
 DE Chicken pp60 c-src protein.  
 KW Endothelial; tyrosine kinase protein; pp60 c-src; ss.  
 OS Gallus gallus.  
 PN WO9314193-A.  
 PD 22-JUL-1993.  
 PF 05-JAN-1993; US004445.  
 PR 06-JAN-1992; US-820011.  
 PA (UYIA ) UNIV YALE.  
 PI Bell L, Luthringer DJ, Madri JA, Warren SL;  
 DR WPI; 93-243209/30.  
 DR P-PSDB; R39705.  
 PT Genetically engineered endothelial cells - which exhibit enhanced  
 PT cell migration, urokinase-type plasminogen activator activity,  
 PT and reduced mononuclear cell adhesion and fibronectin prodn  
 PS Disclosure; Page 64-66; 91pp; English.  
 CC The DNA encoding a portion or (more preferably) the entire pp60  
 CC c-src polypeptide (Given in Q46687) is used to transform endothelial  
 CC cells. Transformed cells produce increased amounts of pp60 c-src and  
 CC have improved therapeutic properties. They migrate at faster rates  
 CC than non-transformed counterparts; have an enhanced ability to  
 CC inhibit the formation of thrombi and/or dissolve thrombi once they  
 CC have formed and exhibit reduced mononuclear cell adhesion. They can  
 CC also be used to improve the success of surgical procedures such as  
 CC coronary angioplasty, heart bypass surgery, vessel graft and stent  
 CC implantation.  
 SQ Sequence 533 AA;

Query Match 53.0%; Score 997; DB 8; Length 533;  
 Best Local Similarity 54.5%; Pred. No. 4.16e-82;  
 Matches 134; Conservative 40; Mismatches 71; Indels 1; Gaps 1;

Db 269 levklggcgvvmgtwngtrvaiktlkpgnmspeafleaqvmmkklrheklvqlvav 328  
 QY 232 LGRKLGEGYGEVWEGWGLWGLSLPVAIKVKSANKMLTDLAKEIOTLKGLRHERLIRHAV 291  
 Db 329 vse-epiyivtymysgslldflkxgmgkylrlpqlvdmagvasgmayvermnyvhrdl 387  
 QY 292 CSQGEPIVITELMRKGNLQALFGLTPEGRALRLPPLLLGFACQVAEGMSYLEEQRVVHRDL 351  
 Db 388 raanilngenlvkvadfglarliedneytarqakfpikwtapeaalgyftlkdsdws 447  
 QY 352 AARNVLVDGLACKVADFGGLARLKDDIYSPSSSSKIPVKWTAPFAAANYRVFSOKSDYMS 411  
 Db 448 fgilltelctkryvpygmvmnrevldqvgymcpcepeslhdmcqcwrrdpeerp 507  
 QY 412 FGVLVHEVFTYGQCPYEGMTNHETLQOIMRGYLRPFAACPAEVYVLMLECWRSPEERP 471  
 Db 508 tfeylq 513  
 QY 472 SFATLR 477

## RESULT 6

ID R39706 standard; Protein; 536 AA.  
 AC R39706;  
 DT 23-DEC-1993 (first entry)  
 DE Human pp60 c-src protein.  
 KW Endothelial; tyrosine kinase protein; pp60 c-src; ss.  
 OS Homo sapien.  
 PN WO9314193-A.  
 PD 22-JUL-1993.  
 PF 05-JAN-1993; US004445.  
 PR 06-JAN-1992; US-820011.  
 PA (UYIA ) UNIV YALE.  
 PI Bell L, Luthringer DJ, Madri JA, Warren SL;  
 DR WPI; 93-243209/30.  
 DR P-PSDB; R39705.  
 PT Genetically engineered endothelial cells - which exhibit enhanced  
 PT cell migration, urokinase-type plasminogen activator activity,  
 PT and reduced mononuclear cell adhesion and fibronectin prodn  
 PS Disclosure; Page 75-77; 91pp; English.  
 CC The DNA encoding a portion or (more preferably) the entire pp60







```

DT 12-FEB-1992 (first entry)
DE Abelson Related Gene, B transcript.
KW Arg; diagnosis; therapy; tumour; abl proto-oncogene.
OS Homo sapiens.
PN US7559029-A.
PD 22-OCT-1991.
PF 22-OCT-1991; 559029.
PR 30-JUL-1990; US-559029.
PA (USSH ) NAT INST OF HEALTH.
PI Kruh G, Aronson SA, King CR;
DR WPI; 91-353425/48.
DR N-PSDB; Q14937.
PT Novel human gene related to abl proto-oncogene - designated
PT "Abelson Related Gene", arg, useful for tumour diagnosis and
PT therapy
PS Disclosure; Fig 5D; 40pp; English.
CC The human gene encoding this protein is closely related to but
CC distinct from the abl proto-oncogene and is a member of the tyrosine
CC kinase encoding family of genes. Arg is expressed as two transcripts.
CC By analogy with c-abl, the alternative 5' arg sequences have been
CC designated A (Q14936) and B and it is assumed that they are joined
CC to the arg second exon.
CC The amino acid sequence is represented as found in the specification.
SQ Sequence 1182 AA;

Query Match 41.7%; Score 785; DB 3; Length 1182;
Best Local Similarity 43.5%; Pred. No. 9.40e-62;
Matches 107; Conservative 49; Mismatches 89; Indels 1; Gaps 1;

Db 288 itmkkhlgghyrevygvkkyistvavtkltkedmeveefikeavmkeikhpnlvql 347
Qy 230 FALGRKLGEYFGEVWEGUWLG-SUPVAIKVKSANKMLTDLAKEIQTLKGRHERLIRL 288
Db 348 lalctleppfyivtymprnildylrqcnrekvtavvilymatgissameylekknfih 407
Qy 299 HAVCSGGEPPVIIVTELMRKNLQAFLGTEGRALRPLLLGFACQVAEGMSYLEEQRVYH 348
Db 408 rllaapnclvgenhvkvadflslrmltadtatthagakfipkwtapeslayntfsiksd 467
Qy 349 RDLAARNVLVDGLACKVADFGALRLKDDIYSPSSSKIPVKWTAPEAANYRVFSQKSD 408
Db 468 vvafigvllweiatygmppgldlsqvydllekgyrmedegpcppkvymracwkwsa 527
Qy 409 VMSFGVLLHEVFTYGCQPYEGMTNHETLQOIMRGYLRPRPAACPAEYVYLMLECWRSPE 468
Db 528 drpsfa 533
Qy 469 ERPSFA 474

RESULT 12
ID R71132 standard; Protein; 466 AA.
AC R71132;
DE N-terminal truncated cytoplasmic tyrosine kinase.
KW cytoplasmic; tyrosine kinase; blood; cell differentiation;
KW screening; anticancer agent; SH3; src homology domain.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 7..70
FT /note= "SH3 domain"
FT domain 81..155
FT /note= "SH2 domain"
FT domain 192..438
FT /note= "tyrosine kinase domain"
PN W09506113-A.
PD 02-MAR-1995.
PF 25-AUG-1994; J01411.
PR 25-AUG-1993; JP-210403.
PR 29-MAR-1994; JP-058553.
PA (ASAH ) ASAH KASEI KOGIO KK.
PI Sakano S;
DR WPI; 95-106842/14.

DR N-PSDB; Q84888.
PT Cytoplasmic tyrosine kinase and antibody recognising it - for
PT screening chemical substances for tyrosine kinase inhibitory or
PT activating activity for use as cancer therapy
PS Claim 1; Page 40-42; 58pp; English.
CC A cytoplasmic tyrosine kinase which has enhanced expression in
CC connection with blood cell differentiation has been isolated from the
CC human UT-7 blood cell line. This sequence comprises an N-terminal
CC truncated fragment of the enzyme (residues 42-507 of R71132). The DNA
CC sequences and antibodies raised against the enzyme, are useful for
CC screening agents for inhibiting or activating activity on the tyrosine
CC kinase, for use as anticancer agents.
SQ Sequence 466 AA;

Query Match 40.9%; Score 769; DB 13; Length 466;
Best Local Similarity 47.6%; Pred. No. 3.18e-60;
Matches 120; Conservative 42; Mismatches 81; Indels 9; Gaps 6;

Db 194 ltlgaigegefgavlgqeylgq-kvavknike-dvtaqafldetavmtkmghenlvrl 251
Qy 230 FALGRKLGEYFGEVWEGUWLGSLPVAIKVKSANKMLTDLAKEIQTLKGRHERLIRLH 289
Db 252 gvllh-ggilyvmehvskgnlvnfltr-gralvntaqlqlqfshvaegmeylekklvh 309
Qy 290 AVCSGGEPPVIIVTELMRKNLQAFLGTEGRAL-RLPPLLGACQVAEGMSYLEEQRVYH 348
Db 310 rllaarnilvsedlvakvsdfglakaerkgl----dssripvkwtapealkhgfkskd 365
Qy 349 RDLAARNVLVDGLACKVADFGALRLKDDIYSPSSSKIPVKWTAPEAANYRVFSQKSD 408
Db 366 vwsfgvllwevfgyrappkmslkeavseavekgymrpepgcpgpvhvlmscweaepa 425
Qy 409 VMSFGVLLHEVFTYGCQPYEGMTNHETLQOIMRGYLRPRPAACPAEYVYLMLECWRSPE 468
Db 426 rrpffkkaekl 437
Qy 469 ERPSFATLEKL 480

RESULT 13
ID R84181 standard; Protein; 507 AA.
AC R84181;
DE 26-MAR-1996 (first entry)
DE Megakaryocyte kinase MKK1.
KW Megakaryocyte kinase-1; MKK1; cytoplasmic tyrosine kinase;
KW cellular signal transduction; leukaemia; thrombocytopenia.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 48..111
FT /label= SH3_domain
FT domain 122..196
FT /label= SH2_domain
FT domain 233..478
FT /label= Catalytic_domain
PN W09529185-A1.
PD 02-NOV-1995.
PF 24-APR-1995; U05008.
PR 22-APR-1994; US-232545.
PR 21-APR-1995; US-426509.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (SUGEN-) SUGEN INC.
PI Gishizky M, Sures I, Ullrich A;
DR WPI; 95-382959/49.
DR N-PSDB; T00616.
PT New poly-nucleotide(s) encoding megakaryocyte tyrosine kinase(s) -
PT used to develop prods. for the treatment and diagnosis of kinase
PT related signal transduction abnormalities.
PS Claim 15; Fig 1A-C; 82pp; English.
CC Human megakaryocyte kinase MKK1 (R84181) is a 58 kDa cytosolic
CC tyrosine kinase showing 54% homology with csk. It appears to play
CC a regulatory role in the growth and differentiation of
CC megakaryocytes and perhaps neural tissues. Recombinant MKK1 can be
CC produced in host cells by expression of encoding cDNA (T00616), and
```

Db 293 gvilh-gglyivmehvskgnlvnflrtr-gralvntaqllqfslhvaegmyleskklvh 350

QY 353 ARNVLDDGLACKVADFGGLARLLKDDIYSPSSSKIPVKWTAPEAANYRVFSQKSDVWSE

Db 551 gvlmweiftegrmpfekntnyevvmtvtrghrlhrpklatkyllyevmlrcwgerpegrps 610  
QY 413 GVLHEVFTYGOCPYEGWTHETLQQIMRGYRUPRPAACPAEVYVLMLECNRSPEERPS 472  
Db 611 fedl 614  
QY 473 FATL 476

Search completed: Thu May 20 13:03:22 1999  
Job time : 174 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 13:04:25 1999; MasPar time 17.36 Seconds

Tabular output not generated. 541.572 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (230-480) from US09099053.pap (5 of 6)  
Perfect Score: 1882  
Sequence: 1 FALGRKLGEGYGEVWGLW.....ECWRSSPERPSFATLREKL 251

Scoring table: PAM 150  
Gap 11

Searched: 116695 seqs, 37453910 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 46.462; Variance 96.799; scale 0.480

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1690	89.8	496	2	A56040 protein-tyrosine-kinase	0.00e+00
2	1670	88.7	496	2	I56322 srm - mouse	5.06e-302
3	1097	58.3	506	2	S24553 protein-tyrosine kinase	1.85e-186
4	1072	57.0	334	2	S24552 protein-tyrosine kinase	1.85e-181
5	1049	55.7	362	2	S24551 protein-tyrosine kinase	7.26e-177
6	1049	55.7	505	2	S24550 protein-tyrosine kinase	7.26e-177
7	1033	54.9	451	2	S49016 protein-tyrosine kinase	1.13e-173
8	1019	54.1	505	2	I38396 protein-tyrosine kinase	7.03e-171
9	1015	53.9	537	2	I51592 p59(Mfn) - xiphophor	4.41e-170
10	1008	53.6	529	1	TVHUPR protein-tyrosine kinase	1.10e-168
11	1006	53.5	532	2	S34104 protein-tyrosine kinase	2.74e-168
12	1006	53.5	532	2	B34104 protein-tyrosine kinase	2.74e-168
13	1007	53.5	534	2	A44931 protein-tyrosine kinase	1.73e-168
14	1004	53.3	537	2	A43866 protein-tyrosine kinase	6.86e-168
15	1002	53.2	512	2	I49552 protein-tyrosine kinase	1.72e-167
16	1000	53.1	537	1	TVHUSY protein-tyrosine kinase	4.30e-167
17	1000	53.1	542	2	A49114 protein-tyrosine kinase	4.30e-167
18	1000	53.1	568	1	TVFV51 protein-tyrosine kinase	4.30e-167
19	997	53.0	533	1	TVCHS protein-tyrosine kinase	1.70e-166
20	993	52.8	526	2	S20808 protein-tyrosine kinase	1.07e-165
21	993	52.8	534	2	S33568 protein-tyrosine kinase	1.07e-165
22	993	52.8	542	1	TVHUCS protein-tyrosine kinase	1.07e-165
23	992	52.7	507	2	A39939 protein-tyrosine kinase	1.69e-165

24	991	52.7	523	1	TVFVMT	protein-tyrosine kinase	2.67e-165
25	992	52.7	539	2	B49114	protein-tyrosine kinase	1.69e-165
26	991	52.7	541	2	A43610	protein-tyrosine kinase	2.67e-165
27	992	52.7	557	1	TVFVS2	protein-tyrosine kinase	1.69e-165
28	992	52.7	587	1	TVFVPR	protein-tyrosine kinase	1.69e-165
29	989	52.6	509	1	TVHAST	protein-tyrosine kinase	6.68e-165
30	989	52.6	541	1	TVCHYS	protein-tyrosine kinase	6.68e-165
31	987	52.4	526	1	OKFVYR	protein-tyrosine kinase	1.67e-164
32	986	52.4	526	2	S15582	protein-tyrosine kinase	2.64e-164
33	986	52.4	526	1	TVFVR	protein-tyrosine kinase	2.64e-164
34	986	52.4	526	1	TVFVGR	protein-tyrosine kinase	2.64e-164
35	983	52.2	528	1	TVFV9	protein-tyrosine kinase	1.04e-163
36	982	52.2	544	2	I51593	protein-tyrosine kinase	1.65e-163
37	981	52.1	526	1	TVFV60	protein-tyrosine kinase	2.61e-163
38	979	52.0	505	2	I37206	protein-tyrosine kinase	6.54e-163
39	976	51.9	509	2	I48845	p56-tck - mouse	2.58e-162
40	976	51.9	541	2	S31645	protein-tyrosine kinase	2.58e-162
41	977	51.9	543	1	TVHUY5	protein-tyrosine kinase	1.63e-162
42	975	51.8	505	1	TVHUC	protein-tyrosine kinase	4.09e-162
43	975	51.8	509	2	A23639	protein-tyrosine kinase	4.09e-162
44	974	51.8	537	2	A45501	protein-tyrosine kinase	6.46e-162
45	973	51.7	536	2	S33569	protein-tyrosine kinase	1.02e-161

## ALIGNMENTS

RESULT 1

ENTRY protein-tyrosine kinase (EC 2.7.1.112) Srm, nonreceptor type  
TITLE  
ORGANISM A56040 #type complete  
DATE 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 12-Jun-1998  
ACCESSIONS A56040  
REFERENCE A56040  
#authors Kohmura, N.; Yagi, T.; Tomooka, Y.; Oyanagi, M.; Kominami, R.; Takeda, N.; Chiba, J.; Ikawa, Y.; Aizawa, S.  
#journal Mol. Cell. Biol. (1994) 14:6915-6925  
#title A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption.  
#accession A56040  
#status preliminary  
#molecule\_type mRNA  
#residues 1-496 #label KOH  
#cross-references GB:D26186; NID:g529072; PID:d1005873; PID:g529073

GENETICS

#map\_position 2  
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology; SH3 homology  
KEYWORDS ATP; phosphotransferase

FEATURE

52-111 #domain SH3 homology #label SH3\  
232-491 #domain protein kinase homology #label KIN\  
240-248 #region protein kinase ATP-binding motif  
SUMMARY #length 496 #molecular-weight 55593 #checksum 301

Query Match 89.8%; Score 1690; DB 2; Length 496;  
Best Local Similarity 85.7%; Pred. No. 0.00e+00;  
Matches 215; Conservative 24; Mismatches 12; Indels 0; Gaps 0;

Db	234	FVGRKLGEGFGEVWGLGSLPVAIVKTSADMDLADLTKEALKSLRHLRLRH	293
Qy	230	FALGRKLGEGFGEVWGLGSLPVAIVKTSANKKLTDLAKEITLGLRHLRLRH	289
Db	294	AICSLGEPVYIVTELGMKGNLQVYLGSGKALSPLHLGFCACVAEGMSYLEERVVHR	353
Qy	290	AVCSGGEVYIVTELMRKGNLQAFGLTPEGRALRPLPLGFCACVAEGMSYLEERVVHR	349
Db	354	DLARNVLVGGDLCKVADFGIARLLKDDVYSPSSGSIPIPKWTAPAEANRVFSQKSDV	413
Qy	350	DLAARNVLVGGDLCKVADFGIARLLKDDVYSPSSGSIPIPKWTAPAEANRVFSQKSDV	409
Db	414	WSFGILLVEVFTYGCPCYEGMTNHTLQISRGYRLPRPACPAEYVYVLMVCEWKGSPPE	473



```
QY 232 LGRKLGEYFGEVWGLWGLSLPVAIKSANKMLTDLAKEIOTLKLGRHERLIRLHAV 291
Db 128 CTIGEPYIVITELMKYGSMLBYLKHGEKNTLHOMVDMQAQISGWYLEAHSVIHRDL 187
QY 292 CSGGEPYIVITELMRKGNLQAFLTGTEGRALRLPPLLGACQVAGMSYLEEQRVVHRDL 351
Db 188 AARNILVGEHGVCKVADFGARLVKEDLYINREGTKFPIKWTAPAEALYNRETIKSDVMS 247
QY 352 AARNVLDDGLACKVADFGARLVKEDLYINREGTKFPIKWTAPAEALYNRETIKSDVMS 411
Db 248 FGVLISETVTHGRMPYGPMTNRQVLEAVDRGYRMPCEGCPDPLKYIKMLSCWKHPEDDRP 307
QY 412 FGVLLHEVFTYGCQYEGTWNHETLQIMRGYRLPRPRACPAEVYVLMLECWRSPEERP 471
Db 308 TFSCLKNL 316
QY 472 SFATLREKL 480

RESULT 5
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) 2 - freshwater sponge
ALTERNATE_NAMES (Spongilla lacustris) (fragment)
ORGANISM src-type tyrosine kinase 2
DATE #formal_name Spongilla lacustris
07-May-1993 #sequence_revision 07-May-1993 #text_change
ACCESSIONS S24551
REFERENCE S24550
#authors Raulf, F.
#submission submitted to the EMBL Data Library, September 1991
#accession S24551
#molecule_type mRNA
#residues 1-362 #label RAU
#cross-references EMBL:X61602; NID:g10151; PID:g10152
GENETICS
#gene srk2
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
KEYWORDS ATP; phosphotransferase; tyrosine-specific protein kinase
FEATURE
1-70 #domain SH2 homology (fragment) #label SH2\
93-351 #domain protein kinase homology #label KIN\
101-109 #region protein kinase ATP-binding motif\
123 #active_site Lys #status predicted
SUMMARY #length 362 #checksum 2776

Query Match 55.7%; Score 1049; DB 2; Length 362;
Best Local Similarity 53.8%; Pred. No. 7.26e-177;
Matches 135; Conservative 50; Mismatches 65; Indels ---T; Gaps 1;

Db 95 ITLIRKLKGAGOFGEVYQGLWNNSTPVAVTKLKGTMQPAFLAEQAQIMKLRHHPKLIQY 154
QY 230 FALGRKLGEGYFGEVWGLWGLSLPVAIKSANKMLTDLAKEIOTLKLGRHERLIRLHV 289
Db 155 AVCTQGEVYIITELMSKGLDLYDLOGEAG-ALKPLQLIDMAAQAAGMAYLELHNYIHR 213
QY 290 AVCSGGEVYIVITELMRKGNLQAFLTGTEGRALRLPPLLGACQVAGMSYLEEQRVVHR 349
Db 214 DLARNILVGNNTCKVADFGARLVSDYNDNATGAKFPIKWTAPAEALFNRFISKDV 273
QY 350 DLARNVLDDGLACKVADFGARLVKEDLYINREGTKFPIKWTAPAEALYNRETIKSDV 409
Db 274 WSGFGLITELTYGRIPYGPMSNAEVLNLDKGYRMPCTVPTPESLYQIMLDCWRNAD 333
QY 410 WSGFGLLHEVFTYGCQYEGTWNHETLQIMRGYRLPRPRACPAEVYVLMLECWRSPEE 469
Db 334 RPTFEALQWL 344
QY 470 RPSATLREKL 480

RESULT 6
```

```
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) 1 - freshwater sponge
ALTERNATE_NAMES (Spongilla lacustris)
ORGANISM src-type tyrosine kinase 1
DATE #formal_name Spongilla lacustris
07-May-1993 #sequence_revision 07-May-1993 #text_change
ACCESSIONS S24550
REFERENCE S24550
#authors Raulf, F.
#submission submitted to the EMBL Data Library, September 1991
#accession S24550
#molecule_type mRNA
#residues 1-505 #label RAU
#cross-references EMBL:X61601; NID:g10149; PID:g10150
GENETICS
#gene srk1
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
KEYWORDS ATP; phosphotransferase; tyrosine-specific protein kinase
FEATURE
61-111 #domain SH3 homology #label SH3\
122-214 #domain SH2 homology #label SH2\
238-496 #domain protein kinase homology #label KIN\
246-254 #region protein kinase ATP-binding motif\
268 #active_site Lys #status predicted
SUMMARY #length 505 #molecular-weight 57693 #checksum 3389

Query Match 55.7%; Score 1049; DB 2; Length 505;
Best Local Similarity 53.7%; Pred. No. 7.26e-177;
Matches 132; Conservative 49; Mismatches 64; Indels 1; Gaps 1;

Db 242 LLRRLGAGOFGEVWGLWNNSTPVAVTKLPGTMSVEEFLOEASIMKLRHHPKLIQYAV 301
QY 232 LGRKLGEYFGEVWGLWGLSLPVAIKSANKMLTDLAKEIOTLKLGRHERLIRLHAV 291
Db 302 CTKEPPIYIVITELMKYGSMLBYLKHGEKNTLHOMVDMQAQISGWYLEAHSVIHRDL 360
QY 292 CSGGEPYIVITELMRKGNLQAFLTGTEGRALRLPPLLGACQVAGMSYLEEQRVVHRDL 351
Db 361 AARNILVGEHGVCKVADFGARLVDEEYIYEATGAKFPIKWTAPAEALYNRETIKSDVMS 420
QY 352 AARNVLDDGLACKVADFGARLVKEDLYINREGTKFPIKWTAPAEALYNRETIKSDVMS 411
Db 421 FGVLYEITTYGRFPYGPMTNPEVLEKIQNVYRMPCPANCPKQFHDIMLDCWRNAD 480
QY 412 FGVLLHEVFTYGCQYEGTWNHETLQIMRGYRLPRPRACPAEVYVLMLECWRSPEERP 471
Db 481 TPTETLQ 486
QY 472 SFATLRL 477

RESULT 7
ENTRY #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) brk - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 07-May-1995 #sequence_revision 21-Jul-1995 #text_change
ACCESSIONS S49016
REFERENCE S49016
#authors Mitchell, P.J.; Barker, K.T.; Martindale, J.E.; Kamalati, T.;
Lowe, P.N.; Page, M.J.; Gusterson, B.A.; Crompton, M.R.
#journal Oncogene (1994) 9:2383-2390
#title Cloning and characterisation of cDNAs encoding a novel
non-receptor tyrosine kinase, brk, expressed in human
breast tumours.
#accession S49016
#status preliminary
#molecule_type mRNA
#residues 1-451 #label MIT
#cross-references EMBL:X78549; NID:g515025; PID:g515026
GENETICS
```

```
#gene GDB:BRK
##cross-references GDB:378058
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
KEYWORDS ATP; phosphotransferase; tyrosine-specific protein kinase
FEATURE
15-67 #domain SH3 homology #label SH3\
78-170 #domain SH2 homology #label SH2\
189-448 #domain protein kinase homology #label KIN\
197-205 #region protein kinase ATP-binding motif
SUMMARY #length 451 #molecular-weight 51834 #checksum 5817

Query Match 54.9%; Score 1033; DB 2; Length 451;
Best Local Similarity 55.6%; Pred. No. 1.13e-173;
Matches 140; Conservative 42; Mismatches 68; Indels 2; Gaps 2;

Db 191 FTLCRLKSGYGEVWEGWGLWGLSLPVAIKVKSANKMLTD-LAKEIOTLGLRHERLRL 250
QY 230 FALGRKLGEYGEVWEGWGLWGLSLPVAIKVKSANKMLTD-LAKEIOTLGLRHERLRL 288

Db 251 YAVVSGDPVYIITELMRKGNLQAFGLTPEGRALRPPLLLGFCQVAGMSYLEEQRVVH 310
QY 289 HAVCSGGEPIVITELMRKGNLQAFGLTPEGRALRPPLLLGFCQVAGMSYLEEQRVVH 348

Db 311 RLAAARNILVGNLTCKVDFGLARLIKEDVY-LSDHNPYKWTAPALSRGHYSTKSD 369
QY 349 RLAAARNVLVDGLACKVADFGARLLKDDIYSPSSSSKIPVKWTAPAEANRYRVFSQSD 408

Db 370 VNSFGILLHEMSRGQVPYPGSNHAEFLVDAGYRMPCEPSPVHKMLTWCARDPE 429
QY 409 VNSFGVLLHEVFTYGCPEYEGMTNHTLQOIMRGYRLPRPAACPAEYVYLMLECWRSPE 468

Db 430 QRPCFKALRRL 441
QY 469 ERPSFATLREKL 480

RESULT 8
ENTRY I38396 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) FRK - human
ALTERNATE_NAMES FYN-related kinase (FRK)
ORGANISM Homo sapiens #common_name man
DATE 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change
09-Apr-1998
ACCESSIONS I38396
REFERENCE I38396
#authors Lee, J.; Wang, Z.; Luoh, S.M.; Wood, W.I.; Scadden, D.T.
#journal Gene (1994) 138:247-251
#title Cloning of FRK, a novel intracellular SRC-like tyrosine kinase-encoding gene.
#cross-references MUID:94171047
#accession I38396
#status preliminary
#molecule_type mRNA
#residues 1-505 #label RES
##cross-references EMBL:U00803; NID:g392887; PID:g392888
GENETICS
#gene GDB:FRK
##cross-references GDB:355675
#map_position 4q35-4q35
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
ATP; phosphotransferase
KEYWORDS
49-105 #domain SH3 homology #label SH3\
232-494 #domain protein kinase homology #label KIN\
240-248 #region protein kinase ATP-binding motif
SUMMARY #length 505 #molecular-weight 58254 #checksum 9379

Query Match 54.1%; Score 1019; DB 2; Length 505;
Best Local Similarity 53.6%; Pred. No. 7.03e-171;
Matches 135; Conservative 43; Mismatches 71; Indels 3; Gaps 2;
```

```
Db 236 LKRLSGQGEVWEGWGLWNNTPVAVKTLKPGSDPNDFLEAQIMKNLRHPKLIQIYAV 295
QY 232 LGRKLGEYGEVWEGWGLWGLSLPVAIKVKSANKMLTDLAKEIOTLGLRHERLRLHAV 291

Db 296 CTLEDPYIITELMRHGSLOEYQLQNDTGSKIHLTQQVDMAAQVAGSMAYLESRYIHRDL 355
QY 292 CSGGEPIVITELMRKGNLQAFGLTPEGRALRPPLLLGFCQVAGMSYLEEQRVVHRDL 351

Db 356 AARNVLGEHNIYKVDAGFLARVFKVDNEDIYESRHIKLPVKWTAPEAIRSNFISKSD 415
QY 352 AARNVLDDGLACKVADFGARLLK-D-DIYSPSSSSKIPVKWTAPEAANRYRVFSQSD 408

Db 416 VNSFGILLYEITYGKMPYSGMTGAQVIQMLAQNRYLPQSPNQPFYIMLECNWAEKP 475
QY 409 VNSFGVLLHEVFTYGCPEYEGMTNHTLQOIMRGYRLPRPAACPAEYVYLMLECWRSPE 468

Db 476 ERPTFETLIRWKL 487
QY 469 ERPSFATLREKL 480

RESULT 9
ENTRY I51592 #type complete
TITLE p59(Xifn) - xiphophorus helleri
ORGANISM #formal_name Xiphophorus helleri
DATE 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change
10-Jul-1998
ACCESSIONS I51592
REFERENCE I51592
#authors Hannig, G.; Ottillie, S.; Scharlt, M.
#journal Oncogene (1991) 6:361-369
#title Conservation of structure and expression of the c-yes and fyn genes in lower vertebrates.
#cross-references MUID:91187435
#accession I51592
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-537 #label HAN
##cross-references EMBL:X54971; NID:g64481; PID:g64482
GENETICS
#gene Xifn
CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
FEATURE
89-138 #domain SH3 homology #label SH3\
269-527 #domain protein kinase homology #label KIN
SUMMARY #length 537 #molecular-weight 60447 #checksum 621

Query Match 53.9%; Score 1015; DB 2; Length 537;
Best Local Similarity 53.7%; Pred. No. 4.41e-170;
Matches 132; Conservative 45; Mismatches 68; Indels 1; Gaps 1;

Db 273 LKRLSGQGEVWMTWNTTAVKTLKPGTMSPEFLEEAQIMKLRHDKLVQIYAV 332
QY 232 LGRKLGEYGEVWEGWGLWGLSLPVAIKVKSANKMLTDLAKEIOTLGLRHERLRLHAV 291

Db 333 VSE-EPIYIVTEYMSKSLDLFLKDGGRALKLPNLVDMAAQVAGSMAYIERMNYIHRDL 391
QY 292 CSGGEPIVITELMRKGNLQAFGLTPEGRALRPPLLLGFCQVAGMSYLEEQRVVHRDL 351

Db 392 RSANILVGNLVCKIADFGFLARLIEDNEYTARQGFPIKWTAPEAALYGRFTTKSDVWS 451
QY 352 AARNVLDDGLACKVADFGARLLKDDIYSPSSSSKIPVKWTAPEAANRYRVFSQSDVWS 411

Db 452 FGILLTELVTGKRVYPGMNNREVLEQVERGYRMPQDCPASHLHMLCOWKDKDPERP 511
QY 412 FGVLLEHFTYGCPEYEGMTNHTLQOIMRGYRLPRPAACPAEYVYLMLECWRSPEERP 471

Db 512 TFEYLO 517
QY 472 SFATLR 477
```

```

RESULT 10
ENTRY TVHUPR #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) fgr - human
ALTERNATE_NAMES kinase-related transforming protein (fgr)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Dec-1988 #sequence_revision 30-Sep-1989 #text_change
22-May-1998
ACCESSIONS A27676; A28353; A24842; A45930; S24306
REFERENCE A27676
#authors Katamine, S.; Notario, V.; Rao, C.D.; Miki, T.; Cheah,
M.S.C.; Tronick, S.R.; Robbins, K.C.
#journal Mol. Cell. Biol. (1988) 8:259-266
#title Primary structure of the human fgr proto-oncogene product p55
(c-fgr)
#cross-references MUID:88094395
#accession A27676
##molecule_type mRNA
##residues 1-529 #label REA
##cross-references GB:M19722; GB:J03429; NID:g182573; PID:g182574
REFERENCE A28353
#authors Inoue, K.; Ikawa, S.; Semba, K.; Sukegawa, J.; Yamamoto, T.;
Toyoshima, K.
#journal Oncogene (1987) 1:301-304
#title Isolation and sequencing of cDNA clones homologous to the
v-fgr oncogene from a human B lymphocyte cell line, IM-9.
#cross-references MUID:8826220
#accession A28353
##molecule_type mRNA
##residues 1-143 #label INO
REFERENCE A24842
#authors Nishizawa, M.; Semba, K.; Yoshida, M.C.; Yamamoto, T.;
Sasaki, M.; Toyoshima, K.
#journal Mol. Cell. Biol. (1986) 6:511-517
#title Structure, expression, and chromosomal location of the human
c-fgr gene.
#cross-references MUID:87064334
#accession A24842
##molecule_type DNA
##residues 111-416 #label REB
##cross-references GB:M12724; NID:g182581; PID:g553286
REFERENCE A45930
#authors Brickell, P.M.; Patel, M.
#journal Br. J. Cancer (1988) 58:704-709
#title Structure and expression of c-fgr protooncogene mRNA in
Epstein-Barr virus converted cell lines.
#accession A45930
##molecule_type mRNA
##residues 1-177;524-529 #label BRI
##cross-references GB:M27454
REFERENCE S24306
#authors Patel, M.; Leeyers, S.J.; Brickell, P.M.
#journal Oncogene (1990) 5:201-206
#title Structure of the complete human c-fgr proto-oncogene and
identification of multiple transcriptional start sites.
#cross-references MUID:90206622
#accession S24306
##status translation not shown
##molecule_type DNA
##residues 1-142 #label PAT
##cross-references EMBL:X52207
GENETICS
#gene GDB:FGR
##cross-references GDB:120615; OMIM:164940
#map_position lp36.2-1p36.1
FUNCTION
#description catalyzes the phosphorylation of a peptidyl tyrosine residue
by ATP
CLASSIFICATION
#superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology
KEYWORDS
ATP; autophosphorylation; blocked amino end; lipoprotein;
myristylation; phosphoprotein; phosphotransferase;
proto-oncogene; thiolester bond; transforming protein;
tyrosine-specific protein kinase

```

```

FEATURE
84-133 #domain SH3 homology #label SH3\
144-241 #domain SH2 homology #label SH2\
261-519 #domain protein kinase homology #label KIN\
269-277 #region protein kinase ATP-binding motif\
2 #modified site myristylated amino end (Gly) (in mature
form) #status predicted\
3,6 #binding_site palmitate (Cys) (covalent) #status
predicted\
291 #active_site Lys #status predicted
SUMMARY #length 529 #molecular-weight 59478 #checksum 2467
Query Match 53.68; Score 1008; DB 1; Length 529;
Best Local Similarity 53.68; Pred. No. 1.10e-168;
Matches 133; Conservative 49; Mismatches 65; Indels 1; Gaps 1;
Db 263 ITLRLRLGTGCGDVLGTWNGSKVAVKTLPGTMSPKAFLEEAQVKNLRLDKLVQLY 322
QY 230 FALGRKLGEYFGEVWELGSLPVAIKVKSANMKLTLDAKEIQLKGLRHERLRLH 289
Db 323 AVVSE-EPIYIVTFMCHGSLDLPLKNPEQDLRLPOLVDMAQVAEGMAYMERMYIHR 381
QY 290 AVCSGGEPVIVITELMRKGNLQAFGLTPEGRALRLPLLGFACQVAEGMSYLEEQRVVHR 349
Db 382 DLRAANILVGERACKTADFLARLIKDDYNPCQSGKFFIKWTAPAALEGRITKSDV 441
QY 350 DLARNVLDVDDGLACKVADFGLARLLKDDIYSPSSSKIPKWTAPAAANRVFSQKSDV 409
Db 442 WSGILLTELITGRIPYCGMKNKREVLQVQEGVHMPGCGPASLYEAMQTVRLQPEE 501
QY 410 WSGVLLHEVTTIYGCQPYEGTTHETLQIMRGYLRPPACPAEVYVLMDECRSSPEE 469
Db 502 RPTFEVIQ 509
QY 470 RPSFATLR 477
RESULT 11
ENTRY A34104 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) src 1 - African clawed
frog
ALTERNATE_NAMES kinase-related transforming protein (src); kinase-related
transforming protein (src) 1
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change
20-Mar-1998
ACCESSIONS A34104; I51564
REFERENCE A34104
#authors Steele, R.E.; Unger, T.F.; Mardis, M.J.; Fero, J.B.
#journal J. Biol. Chem. (1989) 264:10649-10653
#title The two Xenopus laevis SRC genes are co-expressed and each
produces functional pp(60src).
#cross-references MUID:89278134
#accession A34104
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-532 #label STE
##cross-references GB:M24704; GB:J04822; NID:g214804; PID:g214805
REFERENCE I51564
#authors Steele, R.E.; Chosn, R.; Ral, B.B.A.; Winokur, S.T.; Unger,
T.F.
#journal Oncogene (1992) 7:2345-2350
#title Structural organization of a src gene from xenopus laevis.
#cross-references MUID:93064714
#accession I51564
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-113 #label ST2
##cross-references GB:M33646; NID:g214808; PID:g214810
GENETICS
#introns 80/1
CLASSIFICATION
#superfamily protein-tyrosine kinase src; protein kinase
homology; SH2 homology; SH3 homology

```



```

FEATURE
  87-136      #domain SH3 homology #label SH3\
  147-244      #domain SH2 homology #label SH2\
  264-522      #domain protein kinase homology #label KIN\
  272-280      #region protein kinase ATP-binding motif
SUMMARY      #length 332 #molecular-weight 59736 #checksum 7595

Query Match      53.5%; Score 1006; DB 2; Length 532;
Best Local Similarity 54.5%; Pred. No. 2,74e-168;
Matches 134; Conservative 40; Mismatches 71; Indels 1; Gaps 1;

Db      268      LEKLGQCGGEVWGTNGTTRTAIKTKLPGTMSPEAPLQEAQVKKLRHEKLVQLYAV      327
Qy      232      LGRKLGEYGEYEWGLTSLPVAIKVTKSANMKLTDLAKEIQTLKGLRHERLRLHAV      291
Db      328      VSE-EPYIIVTEYMSKGLSLDLFLKGMGYRLRLPOLVDMAAQIAGSMAYVEMNVVHRDL      386
Qy      292      CSGGEPYIVITELMRKGNQAFIGTPEGRALRLPPLGLFCQAVGMSYLEBQRVVHRDL      351
Db      387      RAANILVGENLCKVADFGLARLIEDNEYTAHQGAKFPIKWTAPAAALYGRFTIKSDVWS      446
Qy      352      AARNVLVDDGLACKVADFGLARLLKDDIYSPSSSKIPVKWTAPAAANYRVFSQKSDVWS      411
Db      447      FGILLTELTTKRVYPGMVNVREVLQDVERGYMPCPPDCPSLHDLFMQCWRKDPDEEP      506
Qy      412      FGVLLEHVFTYGCPEYGMTNHTLQIQIMRGYLRPRAPCAPEVYVLMLECWSSRSPERP      471
Db      507      TFEYVQ      512
Qy      472      SFATLR      477

```

```

#type complete
protein-tyrosine kinase (EC 2.7.1.112) fyn - mouse
kinase-related transforming protein (fyn)
#formal_name Mus musculus #common_name house mouse
03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change
08-Sep-1997

ACCESSIONS      A44991
REFERENCE       A44991
#authors       Cooke, M.P.; Perlmutter, R.M.
#journal       New Biol. (1989) 1:66-74
#title         Expression of a novel form of the fyn proto-oncogene in
                hematopoietic cells.
#accession     A44991
#status        preliminary
#molecule_type mRNA
#residues      1-534 #label COO
#cross-references GB:M27366; NID:gi93357; PID:g309241
#note          in the authors' translation an additional residue Leu
                was shown after Lys, for residue 459

CLASSIFICATION  #superfamily protein-tyrosine kinase src; protein kinase
                homology; SH2 homology; SH3 homology
KEYWORDS        ATP; autophosphorylation; phosphoprotein; phosphotransferase
                tyrosine-specific protein kinase

FEATURE
89-138          #domain SH3 homology #label SH3\
149-246         #domain SH2 homology #label SH2\
266-524         #domain protein kinase homology #label KIN\
274-282         #region protein kinase ATP-binding motif
                #length 534 #molecular-weight 60057 #checksum 2574

SUMMARY
Query Match     53.5%; Score 1007; DB 2; Length 534;
Best Local Similarity 52.8%; Pred. No. 1.73e-169;
Matches 130; Conservative 46; Mismatches 69; Indels 1; Gaps 1;

Db 270 LEKKGCGFAEYVLGTWNGNTKVAIKTLKPTGMSPEFLEAQIMKLKDKLVQYAV 329
| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Qy 232 LGRKLEGYGEVWEGWLGSLPVAIKVTKSANMKLTLAKETIQIKGRHERLRLHAV 291
| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 330 VSE-EPYIYVTEYMSKSLDLFKDGEGRALKPLNLVDMAAQVAGMAYIERMNYIHRDL 388
| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```



\*\*\*\*\*  
M P S R E L E H (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 13:07:36 1999; Maspar time 12.30 Seconds  
576.996 Million cell updates/sec

Tabular output not generated.

Title: >US-09-099-053-2  
Description: (230-480) from US09099053.pep (5 of 6)  
Perfect Score: 1882  
Sequence: 1 FALGRKLGEGYGEVWGLW.....ECWRSSPEERPSFATREKL 251

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 47.757; Variance 83.764; scale 0.570

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1670	88.7	496	1 SRM_MOUSE	TYROSINE-PROTEIN KINAS	0.00e+00
2	1097	58.3	506	1 SRK4_SPOLA	TYROSINE-PROTEIN KINAS	4.93e-221
3	1072	57.0	334	1 SRK3_SPOLA	TYROSINE-PROTEIN KINAS	4.18e-215
4	1049	55.7	362	1 SRK2_SPOLA	TYROSINE-PROTEIN KINAS	1.17e-209
5	1049	55.7	505	1 SRK1_SPOLA	TYROSINE-PROTEIN KINAS	1.17e-209
6	1019	54.1	505	1 FRK_HUMAN	TYROSINE-PROTEIN KINAS	1.47e-202
7	1015	53.9	536	1 FYN_XIPHE	PROTO-ONCOGENE TYROSIN	1.30e-201
8	1008	53.6	529	1 FGR_HUMAN	PROTO-ONCOGENE TYROSIN	5.86e-200
9	1006	53.5	531	1 SRC1_XENLA	TYROSINE-PROTEIN KINAS	1.74e-199
10	1006	53.5	531	1 SRC1_XENLA	TYROSINE-PROTEIN KINAS	1.74e-199
11	1007	53.5	533	1 FYN_MOUSE	PROTO-ONCOGENE TYROSIN	1.01e-199
12	1004	53.3	536	1 FYN_XENLA	PROTO-ONCOGENE TYROSIN	5.17e-199
13	1000	53.1	536	1 FYN_HUMAN	PROTO-ONCOGENE TYROSIN	4.55e-198
14	1000	53.1	568	1 SRC_AVIS	TYROSINE-PROTEIN KINAS	4.55e-198
15	997	53.0	532	1 SRC_CHICK	PROTO-ONCOGENE TYROSIN	2.33e-197
16	993	52.8	535	1 SRC_HUMAN	PROTO-ONCOGENE TYROSIN	2.05e-196
17	992	52.7	507	1 LCK_CHICK	PROTO-ONCOGENE TYROSIN	3.53e-196
18	991	52.7	523	1 SRC_RSVPA	TYROSINE-PROTEIN KINAS	6.08e-196
19	992	52.7	533	1 FYN_CHICK	PROTO-ONCOGENE TYROSIN	3.53e-196
20	991	52.7	540	1 SRCN_MOUSE	NEURONAL PROTO-ONCOGEN	6.08e-196
21	992	52.7	557	1 SRC_AVIS1	TYROSINE-PROTEIN KINAS	3.53e-196
22	992	52.7	587	1 SRC_AVIS2	TYROSINE-PROTEIN KINAS	3.53e-196
23	989	52.6	509	1 SPK_HYDAT	TYROSINE-PROTEIN KINAS	1.80e-195

24	989	52.6	541	1 YES_CHICK	PROTO-ONCOGENE TYROSIN	1.80e-195
25	987	52.4	526	1 SRC_RSVH1	TYROSINE-PROTEIN KINAS	5.35e-195
26	986	52.4	526	1 SRC_AVISR	TYROSINE-PROTEIN KINAS	9.22e-195
27	986	52.4	526	1 SRC_RSVR	TYROSINE-PROTEIN KINAS	9.22e-195
28	983	52.2	528	1 YES_AVIS	TYROSINE-PROTEIN KINAS	4.71e-194
29	982	52.2	544	1 YES_XIPHE	PROTO-ONCOGENE TYROSIN	8.11e-194
30	979	52.0	504	1 BLK_HUMAN	TYROSINE-PROTEIN KINAS	4.14e-193
31	978	52.0	526	1 SRC_RSVR	TYROSINE-PROTEIN KINAS	7.13e-193
32	976	51.9	541	1 YES_MOUSE	PROTO-ONCOGENE TYROSIN	2.11e-192
33	977	51.9	543	1 YES_HUMAN	PROTO-ONCOGENE TYROSIN	1.23e-192
34	975	51.8	508	1 LCK_MOUSE	PROTO-ONCOGENE TYROSIN	3.64e-192
35	975	51.8	526	1 HCK_HUMAN	TYROSINE-PROTEIN KINAS	3.64e-192
36	974	51.8	537	1 YES_XENLA	PROTO-ONCOGENE TYROSIN	6.26e-192
37	973	51.7	535	1 YRK_CHICK	PROTO-ONCOGENE TYROSIN	1.08e-192
38	969	51.5	539	1 YES_CANVA	PROTO-ONCOGENE TYROSIN	9.47e-191
39	967	51.4	545	1 FGR_FSVGR	TYROSINE-PROTEIN KINAS	2.81e-190
40	966	51.3	508	1 LCK_HUMAN	PROTO-ONCOGENE TYROSIN	4.83e-190
41	959	51.0	503	1 HCK_RAT	TYROSINE-PROTEIN KINAS	2.15e-188
42	960	51.0	524	1 HCK_MOUSE	TYROSINE-PROTEIN KINAS	1.25e-188
43	957	50.9	517	1 FGR_MOUSE	PROTO-ONCOGENE TYROSIN	6.40e-188
44	954	50.7	511	1 LYN_RAT	TYROSINE-PROTEIN KINAS	3.26e-187
45	951	50.5	511	1 LYN_MOUSE	TYROSINE-PROTEIN KINAS	1.66e-186

ALIGNMENTS

RESULT 1  
ID SRM\_MOUSE STANDARD; PRT; 496 AA.  
AC O62270; O62360;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE SRM (EC 2.7.1.112) (PTK70).  
GN SRMS OR SRM.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=THYMUS;  
RX MEDLINE; 97369678.  
RA KAWACHI Y., NAKAUCHI H., OTSUKA F.;  
RT "Isolation of a cDNA encoding a tyrosine kinase expressed in murine skin".  
RL EXP. DERMATOL. 21:533-538(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RX MEDLINE; 95021220.  
RA KOHMURA N., YAGI T., TOMOOKA Y., OYANAGI M., KOMINAMI R., TAKEDA N.,  
RA CHIBA J., IKAWA Y., AIZAWA S.;  
RT "A novel nonreceptor tyrosine kinase, Src, cloning and targeted disruption".  
RL MOL. CELL. BIOL. 14:6915-6925(1994).  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
PROTEIN TYROSINE PHOSPHATE.  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; D49427; G684972; -.  
DR EMBL; D26186; G529073; -.  
DR MGD; MGI:101865; SRMS.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

```
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P11362; 1FGI.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 55 116 SH3.
FT DOMAIN 124 216 SH2.
FT DOMAIN 234 495 PROTEIN KINASE.
FT NP_BIND 240 248 ATP (BY SIMILARITY).
FT BINDING 262 262 ATP (BY SIMILARITY).
FT ACT_SITE 354 354 BY SIMILARITY.
FT MOD_RES 384 384 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 78 78 R -> G (IN REF. 2).
FT CONFLICT 236 238 LRK -> FGR (IN REF. 2).
FT CONFLICT 278 278 N -> I (IN REF. 2).
SQ SEQUENCE 496 AA; 55731 MW; FD44DEF6 CRC32;

Query Match 88.7%; Score 1670; DB 1; Length 496;
Best Local Similarity 84.9%; Pred. No. 0.00e+00;
Matches 213; Conservative 24; Mismatches 14; Indels 0; Gaps 0;

Db 234 FVLRKGLGEGFGEVWGLGSLPVAIVKISADMKLADLTRENEALKSLRHERLRLH 293
QY 230 FALGRKLGEGFGEVWGLGSLPVAIVKISANKMLTDLAKEIQTGLRHERLRLH 289

Db 294 AICSLGEPYIVITELMKGKLNQVYLSGSEKALSLPHLLGFACQVAGMSYLEERRVVR 353
QY 290 AVCSGEPYIVITELMRKGNLQAFGLTPEGRALRPLPGLGFACQVAGMSYLEEQRVVHR 349

Db 354 DLAAARNVLDGDLTKVADFGFLARLLKDDVYSPSSSKIPVKVTAPEAANYRVFSKSDV 413
QY 350 DLAAARNVLDGDLTKVADFGFLARLLKDDVYSPSSSKIPVKVTAPEAANYRVFSKSDV 409

Db 414 WSGFGLLYEVFTYGCQPYEGMTNHTLQISRGYRLPRPACVPAEYVLMVCEWKGSPEE 473
QY 410 WSGFGLLYEVFTYGCQPYEGMTNHTLQIMRGYRLPRPACVPAEYVLMVCEWKGSPEE 469

Db 474 RPTFAILREKL 484
QY 470 RPSFATLREKL 480

RESULT 2
ID SRK4_SPOLA STANDARD; PRT; 506 AA.
AC P42690;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRK4 (EC 2.7.1.112).
GN SRK4.
OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).
OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;
OC HAPLOSLIDERIDA; SPONGILLIDAE; SPONGILLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92334872.
RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;
RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge
Spongilla lacustris.";
RL ONCOGENE 7:1625-1630(1992).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- ALTERNATIVE PRODUCTS: SRK1 AND SRK4 MAY ARISE BY ALTERNATIVE
SPLICING.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
```

```
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61604; G10156; .
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P00523; 2PTK.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 54 116 SH3.
FT DOMAIN 122 214 SH2.
FT NP_BIND 240 248 ATP (BY SIMILARITY).
FT BINDING 268 268 ATP (BY SIMILARITY).
FT ACT_SITE 359 359 BY SIMILARITY.
SQ SEQUENCE 506 AA; 57561 MW; 807A71D0 CRC32;

Query Match 58.3%; Score 1097; DB 1; Length 506;
Best Local Similarity 54.7%; Pred. No. 4.93e-221;
Matches 134; Conservative 50; Mismatches 60; Indels 1; Gaps 1;

Db 242 LLRGLGAGQGFGEVWGLWNGTTSVAVKTLKPGTMSIEEFLEEAASIMQLRHPKLIQIYAV 301
QY 232 LGRKLGEGFGEVWGLGSLPVAIVKISANKMLTDLAKEIQTGLRHERLRLH 291

Db 302 CTKEEPIYIVITELMKHGSLLLEYL-RGDGRSLKLPDLVDMCSQVAGMSYLEEQNYIHRDL 360
QY 292 CSGEPIYIVITELMRKGNLQAFGLTPEGRALRPLPGLGFACQVAGMSYLEEQRVVHRDL 351

Db 361 AARNILVGEHKICKVADFGFLARVIDEIEYKALGKAPPIKWTAPAEAAAYSRFTIKSDVWS 420
QY 352 AARNVLDDGLACKVADFGFLARLLKDDIYSPSSSKIPVKVTAPEAANYRVFSKSDVWS 411

Db 421 FGIVLYEVITYGRFPYPGMTNAQVLEIQIQSYRMPRMGCPKELYAIMDCWRDPAASRP 480
QY 412 FGVLLEHVEFTYGCQPYEGMTNHTLQIMRGYRLPRPACVPAEYVLMVCEWRSPEERP 471

Db 481 TFEETL 485
QY 472 SFATL 476

RESULT 3
ID SRK3_SPOLA STANDARD; PRT; 334 AA.
AC P42689;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRK3 (EC 2.7.1.112) (FRAGMENT).
GN SRK3.
OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).
OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;
OC HAPLOSLIDERIDA; SPONGILLIDAE; SPONGILLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92334872.
RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;
RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge
Spongilla lacustris.";
RL ONCOGENE 7:1625-1630(1992).
```

CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
CC DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X61603; G10154; -  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50001; SH2; 1.  
DR PROSITE; PS50002; SH3; PARTIAL.  
DR PFAM; PF00017; SH2; 1.  
DR HSP; P12931; 1FMK.  
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
KW SH3 DOMAIN; PHOSPHORYLATION.  
FT NON\_TER 1  
FT DOMAIN <1 42 SH2.  
FT DOMAIN 66 321 PROTEIN KINASE.  
FT NP\_BIND 72 80 ATP (BY SIMILARITY).  
FT BINDING 94 94 ATP (BY SIMILARITY).  
FT ACT\_SITE 186 186 BY SIMILARITY.  
SQ SEQUENCE 334 AA; 37880 MW; 7B64FB9B CRC32;

Query Match 57.08; Score 1072; DB 1; Length 334;  
Best Local Similarity 51.48; Pred. No. 4.18e-215;  
Matches 128; Conservative 58; Mismatches 63; Indels 0; Gaps 0;

Db 68 LQRKLGQNGFGEVWAGVNGTAVAVTKLPDPTMEVDFVQEAQVMKKIKHPNLLQYAV 127  
QY 232 LGRKLGEGYFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIOTLGLRHERLIRH 291  
Db 128 CTIGEPYIVTELMKYGSMLEYLKHGEGKNITLHQWVDMQAISGMYTLEAHSYIHRD 187  
QY 292 CSGGEPYIVTELMRKNLQAFGLTPEGRALRLPPLGLFACQVAEGMSYLEEQRVVHRDL 351  
Db 188 AARNILVGEVNCVADFGFLARVIEDIYNPREGTEKPKIKWTAPEAALYNRFTIKSDVS 247  
QY 352 AARNVLDDGLACKVADFGFLARLKKDDIYSPSSSKIPVKWTAPEAANYRVFSQKSDVS 411  
Db 248 FGVLSIEIVTHGRMPYGMNROVLEAVDRGYRMPCEGCPDPLYKIMLSCWKKEPDPRP 307  
QY 412 FGVLLHEVFTYGCYEGMTNHETLQOIMRGYLRPRAPACPAEVYVLMLECWSSPEER 471  
Db 308 TFSKLNLL 316  
QY 472 SFATLREKL 480

RESULT 4  
ID SRK2-SPOLA STANDARD; PRT; 362 AA.  
AC P42688;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE SRK2 (EC 2.7.1.112) (FRAGMENT).  
GN SRK2.  
OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).  
OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;  
OC HAPLOSCLERIDA; SPONGILLIDAE; SPONGILLA.  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE; 92334872.  
RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;  
RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge  
RL Spongilla lacustris.";  
RL ONCOGENE 7:1625-1630(1992).  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
CC DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X61602; G10152; -  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS50001; SH2; 1.  
DR PROSITE; PS50002; SH3; PARTIAL.  
DR PFAM; PF00017; SH2; 1.  
DR HSP; P00523; 2ETK.  
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;  
KW SH3 DOMAIN; PHOSPHORYLATION.  
FT NON\_TER 1  
FT DOMAIN <1 70 SH2.  
FT DOMAIN 95 348 PROTEIN KINASE.  
FT NP\_BIND 101 109 ATP (BY SIMILARITY).  
FT BINDING 123 123 ATP (BY SIMILARITY).  
FT ACT\_SITE 214 214 BY SIMILARITY.  
SQ SEQUENCE 362 AA; 40937 MW; E00C78BE CRC32;

Query Match 55.74; Score 1049; DB 1; Length 362;  
Best Local Similarity 53.84; Pred. No. 1.17e-209;  
Matches 135; Conservative 50; Mismatches 65; Indels 1; Gaps 1;

Db 95 ITLRLKLGAGFGEVYOGVLWNNSTPVAVTKLKGTMQPAFLAEAOIMKKLRHPLQLY 154  
QY 230 FALGRKLGEGYFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIOTLGLRHERLIRH 289  
Db 155 AVCTGQEPYIVTELMKSGSLDYLQGEAG--ALKLPQLIDMAAQVAGMAYLELHNYIHR 213  
QY 290 AVCSGGEPIVITELMRKNLQAFGLTPEGRALRLPPLGLFACQVAEGMSYLEEQRVVHR 349  
Db 214 DLARNILVGNNTCKVADFGFLARLIVSDDYNATEGAKFPKIKWTAPEAALNRFISKDV 273  
QY 350 DLARNVLVDGLACKVADFGFLARLKKDDIYSPSSSKIPVKWTAPEAANYRVFSQKSDV 409  
Db 274 WSGFGLITLVYGRIPYPGMSNAEVLQNLDKGYRMPCPVTTPESLYQIMLDCWRNAD 333  
QY 410 WSGFVLLHEVFTYGCYEGMTNHETLQOIMRGYLRPRAPACPAEVYVLMLECWSSPEE 469  
Db 334 RPTFEALQWL 344  
QY 470 RPSFATLREKL 480

RESULT 5  
ID SRK1-SPOLA STANDARD; PRT; 505 AA.  
AC P42686;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE TYROSINE-PROTEIN KINASE SRK1 (EC 2.7.1.112).  
GN SRK1.

```

OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).
OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;
RN HAPLOSCLERIDA; SPONGILLIDAE; SPONGILLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92334872.
RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;
RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge
  Spongilla lacustris."
RL ONCOGENE 7:1625-1630(1992).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC -!- PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- ALTERNATIVE PRODUCTS: SRK1 AND SRK4 MAY ARISE BY ALTERNATIVE
CC SPLICING.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X61601; G10150;
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSP: P00523; 2PTK.
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 54 116
FT DOMAIN 122 214
FT DOMAIN 240 493
FT NP_BIND 246 254
FT BINDING 268 268
FT ACT_SITE 359 359
FT ACT_SITE 359 359
SQ SEQUENCE 505 AA; 57693 MW; 3AE3DF34 CRC32;

Query Match
Best Local Similarity 53.7%; Score 1049; DB 1; Length 505;
Matches 132; Conservative .....49; Mismatches...64; Indels...1; Caps 1;

Db 242 LRRLLGAGGFGVWGLMNGTTSVAVKTLKPGTMSVEEFLOEASIMKRLRHPKLIQLYAV 301
Qy 232 LGRKLGEYGFGEWGLMGLSPVAKVYKSNMKLTDLAKETQLKGRHERLFLHAV 291
Db 302 CKKEPIVITVTELMYKSLLEYLRGDEG-VLKIEQLVDVAAVQSGMSYLEQOQNYTHRD 360
Qy 292 CSQGEFVITVTELMRKNQIAFLGTPEGRALRLPLLGACQVAGMSYLEQGVVHRDL 351
Db 361 AARNILVGEHGCKVADFLGAVDEIEYEATGAKFKIKWTAPAEAMYNRTIKSDVWS 420
Qy 352 AARNVLDDGLACKVADFLGALLDDIYSPSSSKIPVKWTAPAEAMYNRTIKSDVWS 411
Db 421 FGWLVEITYYGRFPYGMNPNVELEKIQOQNYRMPKPCOFHDIIMDCWDFPASRP 480
Qy 412 FGVLHVEFTYGCQYEGTGNHETLQOQNYRMPKPCPAEVVYVLMLECRSSPEER 471
Db 481 TFEITQ 486
Qy 472 SFATLR 477

```

---

```

RESULT 6
ID FRK_HUMAN STANDARD; PRT; 505 AA.
AC P42685; Q13128;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE FRK (EC 2.7.1.112) (NUCLEAR TYROSINE PROTEIN
  DE KINASE RAK).
GN FRK.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94171047.
RA LEE J., WANG Z., LUOH S.-M., WOOD W.I., SCADDEN D.T.;
RT "Cloning of FRK, a novel human intracellular SRC-like tyrosine
  RT kinase-encoding gene."
RL GENE 138:247-251(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95210168.
RA CANCE W.G., CRAVEN R.J., BERGMAN M., XU L.H., ALITALO K., LIU E.T.;
RT "Rak, a novel nuclear tyrosine kinase expressed in epithelial cells."
RL CELL GROWTH DIFFER. 5:1347-1355(1994).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE: 93293373.
RA CANCE W.G., CRAVEN R.J., WEINER T.M., LIU E.T.;
RT "Novel protein kinases expressed in human breast cancer."
RN INT. J. CANCER 54:571-577(1993).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: RESTRICTED TO CELLS LINES DERIVED FROM TISSUES
  CC OF LYMPHOID, BRAIN, BREAST, COLON AND BLADDER ORIGIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
  CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00803; G392888;
DR EMBL: U23222; G732528;
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSP: P00523; 2PTK.
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 42 110
FT DOMAIN 116 208
FT DOMAIN 234 491
FT NP_BIND 240 248
FT BINDING 262 262
FT ACT_SITE 354 354
FT MOD_RES 387 387
FT VARIANT 122 122
FT CONFLICT 115 115
FT CONFLICT 115 115
SQ SEQUENCE 505 AA; 58254 MW; C4226A83 CRC32;

```

DR	HSP: P06241; LAOT.	DR	PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; KIN
KW	ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE; LIPIDPROTEIN.	KW	ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE; LIPIDPROTEIN.
FT	INIT_MET 0 0 BY SIMILARITY.	FT	INIT_MET 0 0 BY SIMILARITY.
FT	LIPID 1 1 MYRISTATE (BY SIMILARITY).	FT	LIPID 1 1 MYRISTATE (BY SIMILARITY).
FT	LIPID 2 2 PALMITATE (BY SIMILARITY).	FT	LIPID 2 2 PALMITATE (BY SIMILARITY).
FT	LIPID 5 5 PALMITATE (BY SIMILARITY).	FT	LIPID 5 5 PALMITATE (BY SIMILARITY).
FT	DOMAIN 81 142 SH3.	FT	DOMAIN 81 142 SH3.
FT	DOMAIN 148 245 SH2.	FT	DOMAIN 148 245 SH2.
FT	DOMAIN 270 523 PROTEIN KINASE.	FT	DOMAIN 270 523 PROTEIN KINASE.
FT	MOD_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).	FT	MOD_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT	NP_BIND 276 284 ATP (BY SIMILARITY).	FT	NP_BIND 276 284 ATP (BY SIMILARITY).
FT	BINDING 298 298 ATP (BY SIMILARITY).	FT	BINDING 298 298 ATP (BY SIMILARITY).
FT	ACT_SITE 389 389 BY SIMILARITY.	FT	ACT_SITE 389 389 BY SIMILARITY.
FT	MOD_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).	FT	MOD_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES 530 530 PHOSPHORYLATION (BY SIMILARITY).	FT	MOD_RES 530 530 PHOSPHORYLATION (BY SIMILARITY).
FT	SEQUENCE 536 AA; 60316 MW; BA03DB12 CR32;	FT	SEQUENCE 536 AA; 60316 MW; BA03DB12 CR32;
Query Match	53.9%; Score 1015; DB 1; Length 536;	Query Match	53.9%; Score 1015; DB 1; Length 536;
Best Local Similarity	53.7%; Pred. No. 1.30e-201;	Best Local Similarity	53.7%; Pred. No. 1.30e-201;
Matches 132; Conservative 45; Mismatches 68; Indels 1; Gaps		Matches 132; Conservative 45; Mismatches 68; Indels 1; Gaps	
Db	272 LIRLNGQFGEVMTGNTTKVAVKLUKPTGSPSEFLEBAQIMKKLRHDKLVQIYAV 331	Db	272 LIRLNGQFGEVMTGNTTKVAVKLUKPTGSPSEFLEBAQIMKKLRHDKLVQIYAV 331
Qy	232 LGRKLGEGYFGEVWGLGSLPVAIKVKISANMKLTDLAKEIQTLKGRHERLIRLHAV 291	Qy	232 LGRKLGEGYFGEVWGLGSLPVAIKVKISANMKLTDLAKEIQTLKGRHERLIRLHAV 291
Db	332 VSE-EPYIIVTEYMSKSLDLFLKQEGRALKPLNVDMAQVAGMAYIERMNYIHRDL 390	Db	332 VSE-EPYIIVTEYMSKSLDLFLKQEGRALKPLNVDMAQVAGMAYIERMNYIHRDL 390
Qy	292 CSGEPYIVTELMRKNGLQAFLLGTEGRALRLPLLGFACQVAGMSYLEEQRVVHRDL 351	Qy	292 CSGEPYIVTELMRKNGLQAFLLGTEGRALRLPLLGFACQVAGMSYLEEQRVVHRDL 351
Db	391 RSANILVGNLVCKIADFGRLARLIEDNEYTAHQGAFFIKPTAPEAALYGRFTIKSDVWS 450	Db	391 RSANILVGNLVCKIADFGRLARLIEDNEYTAHQGAFFIKPTAPEAALYGRFTIKSDVWS 450
Qy	352 AARNVLVDGGLACKVADFGRLRLDDIYSPSSSKIPKWTAPAAANYRVFSQSDVWS 411	Qy	352 AARNVLVDGGLACKVADFGRLRLDDIYSPSSSKIPKWTAPAAANYRVFSQSDVWS 411
Db	451 FGILLTELVTGRVPYPGMNNREVLEQVPERGYMPCPODCPASLHELMLQCKWKDPERP 510	Db	451 FGILLTELVTGRVPYPGMNNREVLEQVPERGYMPCPODCPASLHELMLQCKWKDPERP 510
Qy	412 FGVLLEHVEVTGYCQPEYGMTNHETLQQIMRGYLRPRPAACPAEYVYVLMLECWRSSEPERP 471	Qy	412 FGVLLEHVEVTGYCQPEYGMTNHETLQQIMRGYLRPRPAACPAEYVYVLMLECWRSSEPERP 471
Db	511 TFEYVQ 516	Db	511 TFEYVQ 516
Qy	472 SFATLR 477	Qy	472 SFATLR 477
RESULT 8	STANDARD; PRT; 529 AA.	RESULT 8	STANDARD; PRT; 529 AA.
ID FGR_HUMAN		ID FGR_HUMAN	
AC P09769;		AC P09769;	
DT 01-MAR-1989 (REL. 10, CREATED)		DT 01-MAR-1989 (REL. 10, CREATED)	
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)		DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)	
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)		DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)	
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FGR (EC 2.7.1.112) (P55-FGR)		DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FGR (EC 2.7.1.112) (P55-FGR)	
DE (C-FGR)		DE (C-FGR)	
GN FGR OR SRC2.		GN FGR OR SRC2.	
OS HOMO SAPIENS (HUMAN).		OS HOMO SAPIENS (HUMAN).	
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;	
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.		OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.	
RN [1]		RN [1]	
RN SEQUENCE FROM N.A.		RN SEQUENCE FROM N.A.	
RX MEDLINE; 86094395.		RX MEDLINE; 86094395.	
RA KATAMINE S., NOTARIO V., RAO C.D., MIKI T., CHEAH M.S.C.,		RA KATAMINE S., NOTARIO V., RAO C.D., MIKI T., CHEAH M.S.C.,	
RA TRONICK S.R., ROBBINS K.C.;		RA TRONICK S.R., ROBBINS K.C.;	
RT "Primary structure of the human fgr proto-oncogene product p55c-fgr.";		RT "Primary structure of the human fgr proto-oncogene product p55c-fgr.";	
RN MOL. CELL. BIOL. 8:259-266(1988).		RN MOL. CELL. BIOL. 8:259-266(1988).	
RN [2]		RN [2]	
RN SEQUENCE OF 111-416 FROM N.A.		RN SEQUENCE OF 111-416 FROM N.A.	
RX MEDLINE; 87064334.		RX MEDLINE; 87064334.	
RA NISHIZAWA M., SEMBA K., YOSHIDA M.C., YAMAMOTO T., SASAKI M.,		RA NISHIZAWA M., SEMBA K., YOSHIDA M.C., YAMAMOTO T., SASAKI M.,	
RA TOYOSHIMA K.;		RA TOYOSHIMA K.;	
RT "Structure, expression, and chromosomal location of the human c-fgr		RT "Structure, expression, and chromosomal location of the human c-fgr	
RT gene.";		RT gene.";	
RN MOL. CELL. BIOL. 6:511-517(1986).		RN MOL. CELL. BIOL. 6:511-517(1986).	
RN [3]		RN [3]	
RN SEQUENCE OF 1-143 FROM N.A.		RN SEQUENCE OF 1-143 FROM N.A.	





```

QY 232 LGRKLGEGFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIQTGLRHERLIRHAV 291
Db 327 VSE-EPYIVITEYMSKGLSLDLKRGEMGRYLRPLQPLVDMAAQIASGMAYVERMNVYHRDL 385
QY 292 CSGEPPVIVITELMRKGNLQAFGLTPEGRLRPLPGLGFAQVAGMSYLEEQRVVHRDL 351
Db 386 RAANILVGENLVCKVADFGARLIEDNEYTAQGAFFIKWTAPEAALYGRFTIKSDVWS 445
QY 352 AARNVLVDGGLACKVADFGARLIEDNEYTAQGAFFIKWTAPEAALYGRFTIKSDVWS 411
Db 446 FGILLTELTTKGRVPGVMNREVLDQVERGYRMPCCPDPSLHDLMFQCKWRKDEPERP 505
QY 412 FGVLLHEVFTYGCQPYEGMTNHTLQQIMRGYRLPRPAACPAEVYVLMLECHRSSPEERP 471
Db 506 TFEYIQ 511
QY 472 SFATLR 477

RESULT 10
ID SRC1_XENLA STANDARD; PRT; 531 AA.
AC P13115;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRC-1 (EC 2.7.1.112) (P60-SRC-1).
GN SRC-1.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89278134.
RA STEELE R.E., UNGER T.F., MARDIS M.J., PERO J.B.;
RT "The two Xenopus laevis SRC genes are co-expressed and each produces
RT functional pp60src.";
RL J. BIOL. CHEM. 264:10649-10653(1989).
CC -|- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -|- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -|- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M24704; G214805; -.
DR PIR; A34104; A34104.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P00523; 2PTK.
KW TRANSFERASE; ATP-BINDING; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN.
FT INIT_MET 0 0
FT BY SIMILARITY.
FT LIPID 1 1
FT MYRISTATE (BY SIMILARITY).
FT DOMAIN 79 140
FT SH3.
FT DOMAIN 146 243
FT SH2.
FT DOMAIN 265 518
FT PROTEIN KINASE.
FT NP_BIND 271 279
FT ATP (BY SIMILARITY).
FT BINDING 293 293
FT ATP (BY SIMILARITY).

```

```

FT ACT_SITE 384 384 BY SIMILARITY.
FT MOD_RES 414 414 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 531 AA; 59725 MW; AB0705D0 CRC32;

Query Match 53 5%; Score 1006; DB 1; Length 531;
Best Local Similarity 54.5%; Pred. No. 1.74e-199;
Matches 134; Conservative 40; Mismatches 71; Indels 1; Gaps 1;

Db 267 LELKLGQCGFGEVWMTGNTRVAIKTLPGTMSPEAFLOEAQVNMKLRHEKHLVQYAV 326
QY 232 LGRKLGEGFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIQTGLRHERLIRHAV 291
Db 327 VSE-EPYIVITEYMSKGLSLDLKRGEMGRYLRPLQPLVDMAAQIASGMAYVERMNVYHRDL 385
QY 292 CSGEPPVIVITELMRKGNLQAFGLTPEGRLRPLPGLGFAQVAGMSYLEEQRVVHRDL 351
Db 386 RAANILVGENLVCKVADFGARLIEDNEYTAQGAFFIKWTAPEAALYGRFTIKSDVWS 445
QY 352 AARNVLVDGGLACKVADFGARLIEDNEYTAQGAFFIKWTAPEAALYGRFTIKSDVWS 411
Db 446 FGILLTELTTKGRVPGVMNREVLDQVERGYRMPCCPDPSLHDLMFQCKWRKDEPERP 505
QY 412 FGVLLHEVFTYGCQPYEGMTNHTLQQIMRGYRLPRPAACPAEVYVLMLECHRSSPEERP 471
Db 506 TFEYIQ 511
QY 472 SFATLR 477

RESULT 11
ID FYN_MOUSE STANDARD; PRT; 533 AA.
AC P39688;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN FYN.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91175680.
RA COOKE M.P., PERLMUTTER R.M.;
RT "Expression of a novel form of the fyn proto-oncogene in
RT hematopoietic cells.";
RL NEW BIOL. 1:86-74(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA LEE C., KIM M.G., JEON S.H., PARK D.E., PARK S.D., SEONG R.H.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP PALMITOYLATION.
RX MEDLINE; 94019312.
RA SHENOX-SCARIA A.M., TIMSON L.K., KWONG J., SHAW A.S., LUBLIN D.M.;
RT "Palmitoylation of an amino-terminal cysteine motif of protein
RT tyrosine kinases p56lck and p59fyn mediates interaction with
RT glycosyl-phosphatidylinositol-anchored proteins.";
RN [4]
RP PALMITOYLATION.
RX MEDLINE; 95071286.
RA KOEGL M., ZLATKINE P., LEY S.C., COURTNEIDGE S.A., MAGEE A.I.;
RT "Palmitoylation of multiple Src-family kinases at a homologous N-
RT terminal motif.";
RL BIOCHEM. J. 303:749-753(1994).
RN [5]
RP PALMITOYLATION.
RX MEDLINE; 97345356.
RA WOLVEN A., OKAMURA H., ROSENBLATT Y., RESH M.D.;
RT "Palmitoylation of p59fyn is reversible and sufficient for plasma
RT membrane association.";
RL MOL. BIOL. CELL 8:1159-1173(1997).

```







OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TAKEYA T., HANAFUSA H.;  
 RX MEDLINE; 83155664.  
 RT "Structure and sequence of the cellular gene homologous to the RSV  
 RT src gene and the mechanism for generating the transforming virus."  
 RL CELL 32:881-890(1993).  
 RN [2]  
 RP REVISION TO 525.  
 RA TAKEYA T., HANAFUSA H.;  
 RX MEDLINE; 83155664.  
 RT CELL 34:319-319(1993).  
 RN [3]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RA DORAI T., LEVY J.B., KANG L., BRUGGE J.S., WANG L.H.;  
 RX MEDLINE; 91304409.  
 RT "Analysis of cDNAs of the proto-oncogene c-src: heterogeneity in 5'  
 RT exons and possible mechanism for the genesis of the 3' end of  
 RT v-src."  
 RL MOL. CELL. BIOL. 11:4165-4176(1991).  
 RN [4]  
 RP ATP-BINDING SITE.  
 RA KAMPS M.P., TAYLOR S.S., SEFTON B.M.;  
 RX MEDLINE; 84270751.  
 RT "Direct evidence that oncogenic tyrosine kinases and cyclic AMP-  
 RT dependent protein kinase have homologous ATP-binding sites."  
 RL NATURE 310:589-592(1984).  
 RN [5]  
 RP PHOSPHORYLATION.  
 RA GOULD K.L., WOODGETT J.R., COOPER J.A., BUSS J.E., SHALLOWAY D.,  
 RX MEDLINE; 86028181.  
 RT "Protein kinase C phosphorylates pp60src at a novel site."  
 RL CELL 42:849-857(1985).  
 RN [6]  
 RP PHOSPHORYLATION AT TYR-415.  
 RA SMART J.E., OPPERMANN H., CZERNILOFSKY A.P., PURCHIO A.F.,  
 RX MEDLINE; 82082387.  
 RT "Characterization of sites for tyrosine phosphorylation in the  
 RT transforming protein of Rous sarcoma virus (pp60v-src) and its normal  
 RT cellular homologue (pp60c-src)."  
 RL PROC. NATL. ACAD. SCI. U.S.A. 78:6013-6017(1981).  
 RN [7]  
 RP PHOSPHORYLATION AT TYR-526.  
 RA COOPER J.A., GOULD K.L., CARTWRIGHT C.A., HUNTER T.;  
 RX MEDLINE; 86151652.  
 RT "Tyr527 is phosphorylated in pp60c-src: implications for regulation."  
 RL SCIENCE 231:1431-1434(1986).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 82-532.  
 RA WILLIAMS J.C., WEIJLAND A., GONFLONI S., THOMPSON A.,  
 RX MEDLINE; 98070614.  
 RT "The 2.35 A crystal structure of the inactivated form of chicken Src:  
 RT a dynamic molecule with multiple regulatory interactions."  
 RL J. MOL. BIOL. 274:757-775(1997).  
 RN [9]  
 RP STRUCTURE BY NMR OF 80-139.  
 RA YU H., ROSEN M.K., SCHREIBER S.L.;  
 RX MEDLINE; 93279385.  
 RT "1H and 15N assignments and secondary structure of the Src SH3  
 RT domain."  
 RL FEBS LETT. 324:87-92(1993).  
 RN [10]  
 RP STRUCTURE BY NMR OF 76-139.  
 RA FENG S., CHEN J.K., YU H., SIMON J.A., SCHREIBER S.L.;  
 RX MEDLINE; 95063992.  
 RT "Two binding orientations for peptides to the Src SH3 domain:  
 RT development of a general model for SH3-ligand interactions."  
 RL SCIENCE 266:1241-1247(1994).  
 RN [11]  
 RP FUNCTION: THE FUNCTION OF PP60-C-SRC IS UNKNOWN. IT IS EXPRESSED

CC TO HIGH LEVELS, AND WITH A HIGH DEGREE OF KINASE ACTIVITY, IN  
 CC CERTAIN FULLY DIFFERENTIATED CELLS SUCH AS NEURONS, PLATELETS  
 CC AND MACROPHAGES.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- ENZYME REGULATION: BECOMES ACTIVATED WHEN ITS MAJOR TYROSINE  
 CC PHOSPHORYLATION SITE IS NOT PHOSPHORYLATED. IT CAN ALSO BE  
 CC ACTIVATED BY POINT MUTATIONS AS WELL AS BY TRUNCATIONS AT THE  
 CC C-TERMINAL END OR BY OTHER MUTATIONS.  
 CC -1- POLYOMA VIRUS MIDDLE T ANTIGEN FORMS A COMPLEX WITH PP60-C-SRC.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; V00402; G63354; -  
 CC EMBL; S43604; E75862; -  
 CC EMBL; S43616; E75867; -  
 CC EMBL; S43587; E75861; -  
 CC EMBL; S43609; E75863; -  
 CC EMBL; S43614; E75866; -  
 CC PIR; A00630; TVCHS.  
 CC PDB; 1SRU; 31-MAY-94.  
 CC PDB; 1SRM; 31-MAY-94.  
 CC PDB; 2PTK; 24-DEC-97.  
 CC PDB; 1QWE; 08-MAR-96.  
 CC PDB; 1QWF; 08-MAR-96.  
 CC PDB; 1PRU; 07-FEB-95.  
 CC PDB; 1PRM; 07-FEB-95.  
 CC PDB; 1RLP; 07-FEB-95.  
 CC PDB; 1RLQ; 07-FEB-95.  
 CC PDB; 1NLO; 27-JAN-97.  
 CC PDB; 1NLP; 27-JAN-97.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 CC PROSITE; PS50001; SH2; 1.  
 CC PROSITE; PS50002; SH3; 1.  
 CC PFAM; PF00017; SH2; 1.  
 CC PFAM; PF00018; SH3; 1.  
 CC PFAM; PF00069; PKinase; 1.  
 CC TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; PHOSPHORYLATION;  
 CC TRANSFERASE; ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN;  
 CC 3D-STRUCTURE.  
 CC INIT\_MET 0 0  
 CC LIPID 1 1 MYRISTATE.  
 CC FT DOMAIN 80 141 SH3.  
 CC FT DOMAIN 147 244 SH2.  
 CC FT DOMAIN 266 519 PROTEIN KINASE.  
 CC FT MOD\_RES 11 11 PHOSPHORYLATION (BY PKC).  
 CC FT NP\_BIND 272 280 ATP (BY SIMILARITY).  
 CC FT BINDING 294 294 ATP.  
 CC FT ACT\_SITE 385 385 BY SIMILARITY.  
 CC FT MOD\_RES 415 415 PHOSPHORYLATION (AUTO-).  
 CC FT MOD\_RES 526 526 PHOSPHORYLATION.  
 CC SQ SEQUENCE 532 AA; 59919 MW; 65406BF7 CRC32;  
 CC  
 CC Query Match 53.0%; Score 997; DB 1; Length 532;  
 CC Best Local Similarity 54.5%; Pred. No. 2.33e-197;  
 CC Matches 134; Conservative 40; Mismatches 71; Indels 1; Gaps 1;  
 Db 268 LEVKGCCFCFGEVWMTNGTTRVAIKTLKPCNMSPEAFLOEAQVMKKLRHEKLVQLYAV 327  
 Qy 232 LGRKLGEYFGEVWEGWLGSLPVAIKVYSANMKLTLAKETIQLKGLRHERLRLHAV 291



(TM)

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	1077	57.2	451	5	DSRC41.	4.84e-190
2	1033	54.9	517	4	TYROSINE KINASE.	6.75e-181
3	1025	54.5	517	5	SRC-TYPE PROTEIN TYROS	3.09e-179
4	1021	54.3	451	11	SRC-RELATED INTESTINAL	2.09e-178
5	1012	53.8	506	11	SRC RELATED TYROSINE K	1.54e-176
6	1007	53.5	496	5	F49B2 5 PROTEIN.	1.68e-175
7	1003	53.3	527	13	C-SRC TYROSINE KINASE.	1.13e-174
8	1002	53.2	512	11	B-CELL SRC-HOMOLOGY TY	1.83e-174
9	1002	53.2	512	11	B-CELL SRC-HOMOLOGY TY	1.83e-174
10	1000	53.1	534	4	P59FYN.	4.74e-174
11	1000	53.1	537	11	PROTO-ONCOGENE FYN	4.74e-174
12	993	52.8	456	13	NON-RECEPTOR PROTEIN T	1.34e-172
13	993	52.8	456	11	H-19 PROVIRAL SEQUENCE	1.34e-172
14	992	52.7	533	13	GENE C-SRC PRODUCING P	2.16e-172
15	992	52.7	533	13	C-SRC.	2.16e-172
16	991	52.7	587	14	PROTEIN-TYROSINE KINAS	3.48e-172
17	986	52.4	525	14	P60 SRC.	3.78e-171
18	986	52.4	526	14	TYROSINE-PROTEIN KINAS	3.78e-171
19	983	52.2	812	14	COMPLETE GENOME.	1.59e-170
20	976	51.9	505	4	BLK-PROTEIN TYROSINE K	4.45e-169

## ALIGNMENTS

	RESULT	1
ID	Q94879	
AC	Q94879	
DT	01-FEB-2001	
DDT	01-FEB-2001	
DT	01-JAN-2001	
DE	DSRC41	
OC	DROSOPHILA	
OC	EUKARYOTA	
OC	PTERYGOTES	
OC	DROSOPHILA	
RN	[1]	
RP	SEQUENCE	
RC	STRAINING	
RX	MEDLIN	
RY	TAKAHARA	
RT	"Regul."	
RT	Dsrc41	
RT	GENES	
DR	EMBL;	
DDR	FLYBAS	
DDR	PfAM;	
DDR	Pfam;	
DDR	Pfam;	
SQ	SEQUENCE	

	Query Match	57.2%	Score 1077;	DB 5;	Length 517;
	Best Local Similarity	55.8%;	Pred. No. 4,846-190;		
	Matches	139;	Conservative	46;	Mismatches 62;
				Indels	2;
				Gaps	2;
Db	252	RKLGSGQGVWEGVGLWNNITPVAITKLASGTMDPKDFLAEAGIMKRLRTKLIQIYAVCT	311		
Qy	234	RKLGEYFGEVWEGVGLWLSPLVAVIKSANKMLTDLAKEIQTLKGLRHERLIRLHAVCS	293		
Db	312	VEEPIYITELMKHGSLLLEYLQAIAGKSGSLKMWOTLIDMAAOIAGMAYLESQNYIHRDL	371		
Qy	294	GEPVIYITELMKNGNLQAFI-GTPE-GRALRPELLGFACVAGMSYLEQRVVRHDL	351		
Db	372	AARNVLVDGNITVKIADFGLARLIKEDYEARGVAPFKWTAPEAANYSKFISKSDVMS	431		
Qy	352	AARNVLVDLDGLACKVADFGLARLLDDIYSPSSSSKIPVKWTAPEAANYRVFSQKSDVMS	411		
Db	432	FGILITELTYIGRIYPVPGMTNAEVLITQVEHGYRMPQPNCEPRLEYIMLECKHKKOPMRRP	491		





```
RL ONCOGENE 10:349-357(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HSD: ICR; TISSUE=INTESTINE;
RA MEDLINE: 94268846.
RX SIYANOVA E.Y., SERFAS M.S., MAZO I.A., TYNER A.L.;
RT "Tyrosine kinase gene expression in the mouse small intestine.";
RL ONCOGENE 9:2053-2057(1994).
RN [3]
RP SEQUENCE OF 1-77 FROM N.A.
RC STRAIN=BALE/C;
RA SIYANOVA E.Y.;
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
CC TYROSINE PHOSPHATE.
DR EMBL: U16805; G847795; -.
DR EMBL: AF016545; G2738777; -.
DR MGI: 99683; SIK.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
KW TRANSFERASE.
SQ SEQUENCE 451 AA; 51972 MW; 5A749D95 CRC32;

Query Match 54.38; Score 1021; DB 11; Length 451;
Best Local Similarity 55.28; Pred. No. 2.09e-178;
Matches 139; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

Db 191 FTCLKLIGAGYFGEVEALWKKQGVHVAKVYSRDNLLHQHTFOAEIQAMKLRKHLSL 250
QY 230 FAGRLKGEYFGEVWGLWGLSLPVAIKVKSANM-KLTDLAKEIOTLGLRHERLRL 288
Db 251 YAVATAGDPVYIITELMPKGNLQLLRSDSEKALPILELVDFASQVAGMGCYLESNVH 310
QY 289 HAVCSGGEPIVYITELMRKNLQALFTPEGRALRLPLLGFCQVAGMSYLEEQRVVH 348
Db 311 RLAAARNVLVTENLCKVDFGLARLVKEDIY-LSHEHNVPYKWTAPALSRGHYSIKSD 369
QY 349 RLAAARNVLVDGGLACKVDFGLARLVKEDIY-LSHEHNVPYKWTAPALSRGHYSIKSD 408
Db 370 VWSFGVLLHIFSRGQMPYCGMSNHETFLRVADAGYRMPCLPCPNTHKMLSCWSRDPK 429
QY 409 VWSFGVLLHIFSVYTGQCPYEGMTNHTLQIMRGYRLPRPAACPAEYVYVLMLECRSSPE 468
Db 430 QRCFKDLCEKL 441
QY 469 ERPSFATLREKL 480

RESULT 5
ID Q62662 PRELIMINARY; PRT; 506 AA.
AC Q62662;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SRC RELATED TYROSINE KINASE.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE OF 356-416 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RA SUNITHA I., AVIGAN M.I.;
RL BIOCHIM. BIOPHYS. ACTA, GENE STRUCT. EXPR. 1221:348-352(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RX MEDLINE: 96330334.
RA SUNITHA I., AVIGAN M.I.;
RT "The apical membranes of maturing gut columnar epithelial cells contain the enzymatically active form of a newly identified
```

```
RT fyn-related tyrosine kinase.";
RL ONCOGENE 13:547-559(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SMALL INTESTINE;
RA AVIGAN M.I.;
RL SUBMITTED (MAY-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U09583; G339625; -.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
SQ SEQUENCE 506 AA; 58166 MW; 8A3172C1 CRC32;

Query Match 53.84; Score 1012; DB 11; Length 506;
Best Local Similarity 52.84; Pred. No. 1.54e-176;
Matches 133; Conservative 45; Mismatches 71; Indels 3; Gaps 2;

Db 237 LLKRLGSGQFGEVWGLWNTTPVAVKTLKPGSDPNDPDLREAQIMKSLRHPKLIQLYAV 296
QY 232 LGRKLGEYFGEVWGLWGLSLPVAIKVKSANMKLTDLAKEIOTLGLRHERLRLHAV 291
Db 297 CTLEDDPIYIITELMRHGSLOEVLQNDGSKIRLTQQVDMAQVASSGMAYLESQNVHRDL 356
QY 292 CSGEPVYIITELMRKGNLQALFTPEGRALRLPLLGFCQVAGMSYLEEQRVVHRDL 351
Db 357 AARNVLGEHNIYKVADEGLARVFKVDNEDIYESKHEIKLPVKWTAPAIRTNKFSIKSD 416
QY 352 AARNVLVDGGLACKVDFGLARLVK-D--DIYSPSSSSKIPVKWTAPAAANYRVFSQSD 408
Db 417 VWSFGVLLHIFSVYTGQCPYEGMTNHTLQIMRGYRLPRPAACPAEYVYVLMLECRSSPE 476
QY 409 VWSFGVLLHIFSVYTGQCPYEGMTNHTLQIMRGYRLPRPAACPAEYVYVLMLECRSSPE 468
Db 477 QRPTFEELHWKL 488
QY 469 ERPSFATLREKL 480

RESULT 6
ID O45539 PRELIMINARY; PRT; 496 AA.
AC O45539;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F49B2.5 PROTEIN.
DE F49B2.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA KERSHAW J.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SKALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";
RL NATURE 368:32-38(1994).
DR EMBL: Z81543; E1346996; -.
SQ SEQUENCE 496 AA; 56307 MW; C766F63E CRC32;
```

```
Query Match      53.5%; Score 1007; DB 5; Length 496;
Best Local Similarity 53.2%; Pred. No. 1.68e-175;
Matches 133; Conservative 53; Mismatches 61; Indels 3; Gaps 3;

Db 231 LIRQAGCGFGEVGRWNVNVPVAVKKLKATADPTDFLAAQIMKKLRHPKLLSLYAV 290
QY 232 LGRKLGEGYFGEVGEWGLGSLPVAIKVKSANMKLTLDAKEIQTLLKGRHERLRLHAV 291
Db 291 CTRDEPILVTELMGE-NLLTFLQRR-GROCOMPOLVEISAQVAAGMAYLEEMNTHRLD 348
QY 292 CSGGEPVIVTELMRKNLQALGTPEGRALRPLPGLFACQVAGMSYLEEQRVVHRDL 351
Db 349 AARNILINSLSVKIADFLARILMKENEYEARTGAREPIKWTAPAAANYNRTTKSDVW 408
QY 352 AARNVLVDGLACKVADFLAR-LLKDDIYSPSSSKIPVKWTAPAAANYRVFSQSDVW 410
Db 409 SFGILLTEIVTGRPLPYGMNTNAEVLLQVDAGYRMPGACGVPVTLVDIMQOCWRSDPKR 468
QY 411 SFGVLLHEVFTYGQCPYEGMTNHTLQIMRGYRLRPPAACPAEYVYVLMLECWRSPEER 470
Db 469 PTFETLOWKL 478
QY 471 PSFATLREKL 480

RESULT 7
ID Q91952 PRELIMINARY; PRT; 527 AA.
AC Q91952;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE C-SRC TYROSINE KINASE.
GN XSRC
OS XIPHOPHORUS XIPHIDIUM.
OC EURKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA;
OC CYPRINODONTIFORMES; CYPRINODONTOIDEI; POECILIIDAE; XIPHOPHORUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP.F1, RIO SOTO LA MARINA, MEXICO;
RA RAULF F., SAM S., ROBERTSON S.M., MAEUELER W., SCHARTL M.;
RL SUBMITTED (FEB-1992) TO EMBL/GENBANK/DDJB DATA BANKS.
DR EMBL; X64658; G65305; -.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 527 AA; 59078 MW; 40D366A1 CRC32;

Query Match      53.3%; Score 1003; DB 13; Length 527;
Best Local Similarity 54.1%; Pred. No. 1.13e-174;
Matches 133; Conservative 42; Mismatches 70; Indels 1; Gaps 1;

Db 263 LEVKNQCGFGEVGRWNTGTTQVAIKTLKPGTMSPEAFLOQAQVNMKKLRHEKLYVOLYAV 322
QY 232 LGRKLGEGYFGEVGEWGLGSLPVAIKVKSANMKLTLDAKEIQTLLKGRHERLRLHAV 291
Db 323 VSE-EPIVIVTEFMDQGSLLFELKQYSTMRLRLPOLVDFASQIASGMAYVERMNVVHRDL 381
QY 292 CSGGEPVIVTELMRKNLQALGTPEGRALRPLPGLFACQVAGMSYLEEQRVVHRDL 351
Db 382 RAANILVGNLVCKVADFLARLIEDNEYTARQGAKEPIKWTAPAAALGRETIKSDVWS 441
QY 352 AARNVLVDGLACKVADFLARLLKDDIYSPSSSKIPVKWTAPAAANYRVFSQSDVWS 411
Db 442 FGILLVELATKGRVPGMNVREVLDOVGRYRMPGCPSECPASLHMLNCWRKAEERP 501
QY 412 FGVLLEHVFTYGQCPYEGMTNHTLQIMRGYRLRPPAACPAEYVYVLMLECWRSPEER 471
Db 502 TFEYIQ 507
QY 472 SFATLR 477

Query Match      53.2%; Score 1002; DB 11; Length 512;
Best Local Similarity 52.8%; Pred. No. 1.83e-174;
Matches 133; Conservative 45; Mismatches 71; Indels 3; Gaps 2;

Db 243 LKRLGSGFGEVGEWGLWNNTTPVAVKTLKPGSMDPNDLFLRAQIMKSLRHPKLLIOLYAV 302
QY 232 LGRKLGEGYFGEVGEWGLGSLPVAIKVKSANMKLTLDAKEIQTLLKGRHERLRLHAV 291
Db 303 CTLEDPYITITELMRHSGLSQEYLQNDGGSKIHFIOQVDMAAQVASMAYLESQNYTHRDL 362
QY 292 CSGGEPVIVTELMRKNLQALGTPEGRALRPLPGLFACQVAGMSYLEEQRVVHRDL 351
Db 363 AARNVLVGEHNIYKVADFLARVVDNEDYIESKHEIKLPVKWTAPAIRTNKFSIKSD 422
QY 352 AARNVLVDGLACKVADFLARLLK-D--DIYSPSSSKIPVKWTAPAAANYRVFSQSD 408
Db 423 VMSFGILLYEITYGKMPYSGMTGAQVIOMLSONYRLPOPSNCPOQFYSIMLECNVPEK 482
QY 409 VMSFGVLLHEVFTYGQCPYEGMTNHTLQIMRGYRLRPPAACPAEYVYVLMLECWRSPE 468
Db 483 QRPTFETLHWKL 494
QY 469 ERPSPATLREKL 480

RESULT 9
ID Q61364 PRELIMINARY; PRT; 512 AA.
AC Q61364;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE B-CELL SRC-HOMOLOGY TYROSINE KINASE (PROTEIN TYROSINE KINASE).
GN FRK OR BSK.
OS MUS MUSCULUS (MOUSE).
OC EURKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95137395.
RA OBERG-WELSH C., WELSH M.;
RT "Cloning of BSK, a murine FRK homologue with a specific pattern of
RL tissue distribution.";
RL GENE 152:239-242(1995).
DB EMBL; L36132; G77773; -.
```





QY 352 AARNVLDDGLACKVADFGIARLLKDDIYSPSSSSKIPVKWTAPAAANYRVFSQKSDVWS 411  
Db 448 FGILLTETATKGRVPYPMVNVREVLDQVERGYRMPCEPCESLHDLMCQCWRKDPERP 507  
QY 412 FGVLLEHVFTYGCPEYEGMTNHETLQIMRGYRLPRPAACPAEYVYVLMLECWRSRSPERP 471  
Db 508 TFEYVQ 513  
QY 472 SFATLR 477

RESULT 15  
ID Q90992 PRELIMINARY; PRT; 533 AA.  
AC Q90992;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE C-SRC.  
GN SRC.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA HANAFUSA H.;  
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 83155664.  
RA TAKEYA T., HANAFUSA H.;  
RT "Structure and sequence of the cellular gene homologous to the RSV  
RT src gene and the mechanism for generating the transforming virus.";  
RL CELL 32:881-890(1983).  
DR EMBL; J00844; G901820; .  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PFAM; PF00017; SH2; 1.  
DR PFAM; PF00018; SH3; 1.  
DR PFAM; PF00069; pkinase; 1.  
SQ SEQUENCE 533 AA; 60038 MW; 8E987D6B CRC32;

Query Match 52.78; Score 992; DB 13; Length 533;  
Best Local Similarity 54.18; Pred. No. 2.16e-172;  
Matches 133; Conservative 40; Mismatches 72; Indels 1; Gaps 1;  
Db 269 LEVKLGCGCFGEVVMGTWNGTTRVAIKTLKPGTMSPEAFLOEAQVMMKKLRHEKLVQLYAV 328  
QY 232 LGRKLGEYGEVWEGUWLGSLPVAIKVIKSANKMLTDLAKEIOTLKGLRHERLIRLHAV 291  
Db 329 VSE-EPYIVTEYMSKGLDLFLKGMGKYLRPLQVDMAAQIASGMAYVERMNVVHRDL 387  
QY 292 CSGGEPYIVTELMRKGNLQAFGLTPEGRALRPLPLGFCQVAGMSYLEEQRVVHRDL 351  
Db 388 RAANILVGENLVCKVADFGIARLIEDNEYTAQCAKAPPIKWTAPEAAIYGRFTTKSDVWS 447  
QY 352 AARNVLDDGLACKVADFGIARLLKDDIYSPSSSSKIPVKWTAPAAANYRVFSQKSDVWS 411  
Db 448 FGILLTETATKGRVPYPMVNVREVLDQVERGYRMPCEPCESLHDLMCQCWRKDPERP 507  
QY 412 FGVLLEHVFTYGCPEYEGMTNHETLQIMRGYRLPRPAACPAEYVYVLMLECWRSRSPERP 471  
Db 508 TFEYVQ 513  
QY 472 SFATLR 477

Search completed: Thu May 20 13:12:45 1999  
Job time : 248 secs.

\*\*\*\*\*

WQ5RELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 20 13:15:14 1999; MasPar time 4.53 Seconds  
Tabular output not generated.

Title: >US-09-099-053-2  
Description: (481-488) from US09099053.pep (6 of 6)  
Perfect Score: 68  
Sequence: 1 HAIHRCHP 8  
Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneses35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 15.050; Variance 43.650; scale 0.345

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	100.0	488	39	W89248 Human SAD.	2.32e-01
2	46	67.6	4655	29	W43313 Human kidney calcium	9.75e-01
3	46	67.6	4655	29	W43312 Human placental calci	9.75e-01
4	46	67.6	4655	29	W43311 Human calcium sensor	9.75e-01
5	46	67.6	4655	17	R97210 Human kidney calcium	9.75e-01
6	46	67.6	4655	17	R97208 Human calcium sensor	9.75e-01
7	46	67.6	4655	17	R97211 Human parathyroid cal	9.75e-01
8	46	67.6	4655	17	R97209 Human placental calci	9.75e-01
9	46	67.6	4655	29	W43314 Human parathyroid cal	9.75e-01
10	45	66.2	69	38	W88724 Secreted protein enco	1.26e-02
11	45	66.2	84	37	W4745 Human secreted protei	1.26e-02
12	44	64.7	4544	9	R47861 Alpha 2-macroglobulin	1.63e-02
13	44	64.7	4544	11	R60317 Human alpha-2-MR.	1.63e-02
14	43	63.2	9	22	W10404 Antipholipid anal	2.11e-02
15	43	63.2	9	36	W67128 Antibody ACA-6501 epi	2.11e-02
16	43	63.2	255	20	W01200 Serine protease PfSP2	2.11e-02

17	43	63.2	343	21	W01561 MOC1 protein.	2.11e-02
18	43	63.2	510	7	R34674 Insulinoma-associated	2.11e-02
19	43	63.2	709	19	R99801 CRII-7 nerve protein,	2.11e-02
20	42	61.8	95	36	W74972 Human secreted protein,	2.72e-02
21	42	61.8	331	14	R80066 Human IFNAB-BPI.	2.72e-02
22	42	61.8	372	36	W72164 HSV-2 strain SB5 Cont	2.72e-02
23	42	61.8	433	35	W72012 HSV-2 strain SB5 Cont	2.72e-02
24	42	61.8	433	36	W72144 HSV-2 strain SB5 Cont	2.72e-02
25	42	61.8	542	32	W57332 Glycero1-3-phosphate	2.72e-02
26	42	61.8	542	32	W60260 Klebsiella pneumoniae	2.72e-02
27	42	61.8	595	32	W58390 Human secreted protei	2.72e-02
28	42	61.8	650	13	R71382 Drosophila semaphorin	2.72e-02
29	42	61.8	668	29	W55483 H. pylori ORF 14p108	2.72e-02
30	42	61.8	677	29	W55328 H. pylori ORF hp3e103	2.72e-02
31	42	61.8	1594	5	P81185 Sequence encoded by t	2.72e-02
32	41	60.3	18	3	R15221 Residues 380-397 of o	3.50e-02
33	41	60.3	67	30	W46751 Protein encoded by cl	3.50e-02
34	41	60.3	67	29	W42031 AP224_2 protein.	3.50e-02
35	41	60.3	1931	5	R27649 Human calcium channel	3.50e-02
36	41	60.3	2237	6	R33550 Sequence of the alpha	3.50e-02
37	41	60.3	2237	33	W63142 Human calcium channel	3.50e-02
38	41	60.3	2237	14	R71006 Human neuronal calciu	3.50e-02
39	41	60.3	2237	32	W37878 Human calcium channel	3.50e-02
40	41	60.3	2339	14	R71005 Human neuronal calciu	3.50e-02
41	41	60.3	2339	6	R33549 Sequence of the alpha	3.50e-02
42	41	60.3	2339	33	W63141 Human calcium channel	3.50e-02
43	41	60.3	2466	35	W75999 Intracellular protein	3.50e-02
44	41	60.3	2466	13	R71498 Human protein tyrosin	3.50e-02
45	40	58.8	133	4	R22022 Human thrombomodulin	4.49e-02

ALIGNMENTS

RESULT 1  
ID W89248 standard; Protein; 488 AA.

AC W89248;  
DT 10-MAR-1999 (first entry)

DE Human SAD.

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease.  
OS Homo sapiens.

PN W09849317-A2.

PD 05-NOV-1998.

PF 27-APR-1998; U08439.

PR 23-OCT-1997; US-063595.

PR 28-APR-1997; US-044428.

PR 20-MAY-1997; US-047222.

PR 11-JUN-1997; US-049477.

PR 11-JUN-1997; US-049756.

PR 18-JUN-1997; US-049914.

PA (SUGEN-) SUGEN INC.

PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,

PI Markby D, Onrust S, Peles E, Plozman GD;

DR WPI: 99-009434/01.

DR N-PSDB; V81743.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Claim 2; Page 154-155; 193pp; English.

CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC present sequence represents human SAD. The above proteins, other than  
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
CC substances that modulate their activity (i.e. agonists and antagonists,  
CC including NBP) in vivo or in vitro. These substances are used to treat  
CC or prevent diseases associated with abnormal signal transduction  
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
CC serine/threonine kinase) are used to promote neuronal survival,  
CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
CC diseases. Nucleic acid fragments of the polynucleotides encoding the

CC proteins can be used as probes to identify and clone related sequences;  
 CC to detect protein-encoded RNA; to generate transgenic animals and in  
 CC gene therapy (optionally after mutation). Ab are used to determine the  
 CC proteins.  
 SQ Sequence 488 AA;

Query Match 100.0%; Score 68; DB 39; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 2.32e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 haihrchp 488  
 Qy 481 HAIHRCHP 488

RESULT 2  
 ID W43313 standard; Protein; 4655 AA.  
 AC W43313;  
 DT 08-JUN-1998 (first entry)  
 DE Human kidney calcium sensor protein.  
 KW Calcium sensor; human; glycoprotein 330; gp330; kidney;  
 KW low density lipoprotein; LDL; membranous glomerulonephritis;  
 KW osteoporosis; hyperthyroidism; therapy.  
 OS Homo sapiens.

FT Key Location/Qualifiers  
 FT Peptide 1..25  
 FT Domain 26..4422 /label= Sig\_peptide  
 FT Domain /note= "extracellular domain"  
 FT Domain 4423..4445 /note= "transmembrane domain"  
 FT Domain 4446..4655 /note= "C-terminal domain"  
 FT Misc\_difference 4210 /note= "encoded by MTC"

FN WO9744050-A1.  
 PD 27-NOV-1997.  
 PF 20-MAY-1997; U08682.  
 PR 23-MAY-1996; US-652877.  
 PA (RHON ) RHONE-POULENC RORER PHARM INC.  
 PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,  
 PI Murray EM, Rask L;  
 DR WPI: 98-018219/02.  
 DR N-PSDB: V05996.

PT New human calcium sensor protein and vector encoding inhibitor of gp  
 PT 330 mediated transport of amyloid beta protein - useful to treat  
 PT membranous glomerulonephritis and central nervous system disorders  
 PS Claim 7; Page 145-163; 240pp; English.  
 CC This sequence comprises a human kidney calcium sensor protein  
 CC (hCSP), deduced from the nucleotide-sequence (see V05996) of a  
 CC clone obtained by PCR amplification. Full-length sequences for  
 CC human placenta (see W43312) and parathyroid (see W43314) are also  
 CC claimed. hCSP is considered to be the human homologue of rat  
 CC Heymann nephritis antigen gp330 on the basis of sequence  
 CC similarity. It is proposed that the calcium sensor is a  
 CC universal sensor for recognition of variation in extracellular  
 CC calcium, and plays a key role in calcium regulation via different  
 CC organ systems. hCSP belongs to the low density lipoprotein  
 CC superfamily of glycoproteins. It can be expressed in a claimed  
 CC host cell using a vector comprising hCSP cDNA. Also claimed are a  
 CC method of using hCSP to treat human membranous glomerulonephritis,  
 CC and the use of hCSP to screen for agonists and antagonists of hCSP.  
 CC Such (antagonists are useful in the treatment of conditions  
 CC associated with perturbations in parathyroid hormone levels,  
 CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast  
 CC activity, calcium secretion and calcium ion homeostasis, e.g.  
 CC hyperthyroidism and osteoporosis.  
 SQ Sequence 4655 AA;

Query Match 67.6%; Score 46; DB 29; Length 4655;  
 Best Local Similarity 50.0%; Pred. No. 9.75e+01;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hdvhkscp 268  
 Qy 481 HAIHRCHP 488

RESULT 3  
 ID W43312 standard; Protein; 4655 AA.  
 AC W43312;  
 DT 08-JUN-1998 (first entry)  
 DE Human placental calcium sensor protein.  
 KW Calcium sensor; human; glycoprotein 330; gp330; placenta;  
 KW low density lipoprotein; LDL; membranous glomerulonephritis;  
 KW osteoporosis; hyperthyroidism; therapy.  
 OS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT Peptide 1..25  
 FT Domain 26..4422 /label= Sig\_peptide  
 FT Domain /note= "extracellular domain"  
 FT Domain 4423..4445 /note= "transmembrane domain"  
 FT Domain 4446..4655 /note= "C-terminal domain"  
 FT Misc\_difference 213 /note= "encoded by GAY"  
 FT Misc\_difference 675 /note= "encoded by GTY"  
 FT Misc\_difference 1272 /note= "encoded by CCH"  
 FT Misc\_difference 1287 /note= "encoded by SCA"  
 FT Misc\_difference 4094 /note= "encoded by RAG"  
 FT Misc\_difference 4210 /note= "encoded by MTC"  
 FT Misc\_difference 4371 /note= "encoded by ATY"

FN WO9744050-A1.  
 PD 27-NOV-1997.  
 PF 20-MAY-1997; U08682.  
 PR 23-MAY-1996; US-652877.  
 PA (RHON ) RHONE-POULENC RORER PHARM INC.  
 PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,  
 PI Murray EM, Rask L;  
 DR WPI: 98-018219/02.  
 DR N-PSDB: V05995.  
 PT New human calcium sensor protein and vector encoding inhibitor of gp  
 PT 330 mediated transport of amyloid beta protein - useful to treat  
 PT membranous glomerulonephritis and central nervous system disorders  
 PS Claim 7; Page 112-130; 240pp; English.  
 CC This sequence comprises a human placental calcium sensor protein  
 CC (hCSP), deduced from the nucleotide sequence (see V05995) of a  
 CC clone obtained by PCR amplification. Full-length sequences for  
 CC human kidney (see W43313) and parathyroid (see W43314) are also  
 CC claimed. hCSP is considered to be the human homologue of rat  
 CC Heymann nephritis antigen gp330 on the basis of sequence  
 CC similarity. It is proposed that the calcium sensor is a  
 CC universal sensor for recognition of variation in extracellular  
 CC calcium, and plays a key role in calcium regulation via different  
 CC organ systems. hCSP belongs to the low density lipoprotein  
 CC superfamily of glycoproteins. It can be expressed in a claimed  
 CC host cell using a vector comprising hCSP cDNA. Also claimed are a  
 CC method of using hCSP to treat human membranous glomerulonephritis,  
 CC and the use of hCSP to screen for agonists and antagonists of hCSP.  
 CC Such (antagonists are useful in the treatment of conditions  
 CC associated with perturbations in parathyroid hormone levels,  
 CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast  
 CC activity, calcium secretion and calcium ion homeostasis, e.g.  
 CC hyperthyroidism and osteoporosis.  
 SQ Sequence 4655 AA;

Query Match 67.6%; Score 46; DB 29; Length 4655;  
 Best Local Similarity 50.0%; Pred. No. 9.75e+01;

Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hdvbkscp 268  
| :|:|  
QY 481 HAIHRCHP 488

## RESULT 4

ID W43311 standard; Protein; 4655 AA.  
AC W43311;  
DT 08-JUN-1998 (first entry)  
DE Human calcium sensor protein.  
KW Calcium sensor; human; glycoprotein 330; gp330; placenta; kidney;  
KW low density lipoprotein; LDL; membranous glomerulonephritis;  
KW osteoporosis; hyperthyroidism; therapy.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..25  
FT Domain 26..4422  
FT /note= "extracellular domain"  
FT Domain 4423..4445  
FT /note= "transmembrane domain"  
FT Domain 4446..4655  
FT /note= "C-terminal domain"  
PN W09744050-A1.  
PD 27-NOV-1997.  
PF 20-MAY-1997; U08682.  
PR 23-MAY-1996; US-652877.  
PA (RHON ) RHONE-POULENC RORER PHARM INC.  
PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC,  
PI Murray EM, Rask L;  
DR WPI; 98-018219/02.  
DR N-PSDB: V05994.  
DT New human calcium sensor protein and vector encoding inhibitor of gp  
PT 330 mediated transport of amyloid beta protein - useful to treat  
PT membranous glomerulonephritis and central nervous system disorders  
PS Claim 7; Page 80-98; 240pp; English.  
CC This sequence comprises a human calcium sensor protein (hCSP),  
CC deduced from the nucleotide sequence (see V05994) of clones  
CC isolated from a human placental and kidney libraries. Full-length  
CC sequences (see W43312-14) for human kidney, placenta and  
CC parathyroid are also claimed. hCSP is considered to be the human  
CC homologue of rat Heymann nephritis antigen gp330 on the basis of  
CC sequence similarity. It is proposed that the calcium sensor is a  
CC universal sensor for recognition of variation in extracellular  
CC calcium, and plays a key role in calcium regulation via different  
CC organ systems. hCSP belongs to the low density lipoprotein  
CC superfamily of glycoproteins. It can be expressed in a claimed  
CC host cell using a vector comprising hCSP cDNA. Also claimed are a  
CC method of using hCSP to treat human membranous glomerulonephritis,  
CC and the use of hCSP to screen for agonists and antagonists of hCSP.  
CC Such (ant)agonists are useful in the treatment of conditions  
CC associated with perturbations in parathyroid hormone levels,  
CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast  
CC activity, calcium secretion and calcium ion homeostasis, e.g.  
CC hyperthyroidism and osteoporosis.  
SQ Sequence 4655 AA;

Query Match 67.6%; Score 46; DB 29; Length 4655;  
Best Local Similarity 50.0%; Pred. No. 9.75e+01;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hdvbkscp 268  
| :|:|  
QY 481 HAIHRCHP 488

## RESULT 5

ID R97210 standard; Protein; 4655 AA.  
AC R97210;  
DT 31-JUL-1996 (first entry)  
DE Human kidney calcium sensor protein.

KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;  
KW membranous glomerulonephritis; therapy; ulcer; wound; kidney.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 4210  
FT /note= "unidentified amino acid"  
PN W09615801-A1.  
PD 30-MAY-1996.  
PF 22-NOV-1995; U15203.  
PR 23-NOV-1994; US-344836.  
PR 07-JUN-1995; US-487314.  
PA (RHON ) RHONE-POULENC RORER PHARM INC.  
PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC;  
PI Murray EM, Rask L;  
DR WPI; 96-268334/27.  
DR N-PSDB: T28775.  
PT Isolated human calcium sensor protein - useful develop prods. for  
PT diagnosis and treatment of conditions, such as membranous  
PT glomerulonephritis  
PS Claim 7; Page 143-161; 236pp; English.  
CC 500 kba human calcium sensor proteins (CSP) (R97209-11) were  
CC identified as the products of cDNA clones (T28774-76) respectively  
CC obtd. from human placental, kidney and parathyroid cDNA libraries.  
CC The CSPs show homology to LDL receptor proteins. CSP can be isolated  
CC from parathyroid, placental or kidney tubule cells or obtd. by  
CC expression of the encoding sequences in transformed host cells. It is  
CC useful for identifying potential (ant)agonists and in the diagnosis and  
CC treatment of e.g. membranous glomerulonephritis, and also in the  
CC treatment of wounds and ulcers.  
SQ Sequence 4655 AA;

Query Match 67.6%; Score 46; DB 17; Length 4655;  
Best Local Similarity 50.0%; Pred. No. 9.75e+01;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hdvbkscp 268  
| :|:|  
QY 481 HAIHRCHP 488

## RESULT 6

ID R97208 standard; Protein; 4655 AA.  
AC R97208;  
DT 31-JUL-1996 (first entry)  
DE Human calcium sensor protein.  
KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;  
KW membranous glomerulonephritis; therapy; ulcer; wound; kidney.  
OS Homo sapiens.  
PN W09615801-A1.  
PD 30-MAY-1996.  
PF 22-NOV-1995; U15203.  
PR 23-NOV-1994; US-344836.  
PR 07-JUN-1995; US-487314.  
PA (RHON ) RHONE-POULENC RORER PHARM INC.  
PI Akerstrom G, Crumley GR, Hjalm G, Juhlin C, Morse CC;  
PI Murray EM, Rask L;  
DR WPI; 96-268334/27.  
DR N-PSDB: T28773.  
PT Isolated human calcium sensor protein - useful develop prods. for  
PT diagnosis and treatment of conditions, such as membranous  
PT glomerulonephritis  
PS Claim 7; Page 78-96; 236pp; English.  
CC A 500 kDa human calcium sensor protein (CSP) (R97208) was identified  
CC as the product of cDNA clone (T28773) obtd. from human kidney and  
CC placental cDNA libraries. Full-length sequences of placental, kidney  
CC and parathyroid CSP (see also R97209-11) were also obtd. They show  
CC homology to LDL receptor proteins. CSP can be isolated from  
CC parathyroid, placental or kidney tubule cells or obtd. by expression of  
CC encoding sequences in transformed host cells. It is useful for  
CC identifying potential (ant)agonists and in the diagnosis and treatment  
CC of e.g. membranous glomerulonephritis, and also in the treatment of  
CC wounds and ulcers.  
SQ Sequence 4655 AA;



Query Match 67.6%; Score 46; DB 17; Length 4655;  
 Best Local Similarity 50.0%; Pred. No. 9.75e+01;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hvdhkcsp 268  
 | :|:|  
 QY 481 HAIHRCHP 488

RESULT 7

ID R97211 standard; Protein; 4655 AA.  
 AC R97211;  
 DT 31-JUL-1996 (first entry)  
 DE Human parathyroid calcium sensor protein.  
 KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;  
 KW membranous glomerulonephritis; therapy; ulcer; wound; parathyroid.  
 OS Homo sapiens.  
 PN WO9615801-A1.  
 PD 30-MAY-1996.  
 PF 22-NOV-1995; U15203.  
 PR 23-NOV-1994; US-344836.  
 PR 07-JUN-1995; US-487314.  
 PA (RHON ) RHONE-POULENC RORER PHARM INC.  
 PI Amerstrom G, Crumley GR, Hjaln G, Juhlin C, Morse CC;  
 PI Murray EM, Rask L;  
 DR WPI: 96-268334/27.  
 DR N-PSDB: T28776.  
 PT Isolated human calcium sensor protein - useful develop prods. for  
 PT diagnosis and treatment of conditions, such as membranous  
 PT glomerulonephritis  
 PS Claim 7; Page 110-128; 236pp; English.  
 CC 500 kDa human calcium sensor proteins (CSP) (R97209-11) were  
 CC identified as the products of cDNA clones (T28774-76) respectively  
 CC obtd. from human placental, kidney and parathyroid cDNA libraries.  
 CC The CSPs show homology to LDL receptor proteins. CSP can be isolated  
 CC from parathyroid, placental or kidney tubule cells or obtd. by  
 CC expression of the encoding sequences in transformed host cells. It is  
 CC useful for identifying potential (ant)agonists and in the diagnosis  
 CC and treatment of e.g. membranous glomerulonephritis, and also in the  
 CC treatment of wounds and ulcers.  
 CC Sequence 4655 AA;  
 SQ

Query Match 67.6%; Score 46; DB 17; Length 4655;  
 Best Local Similarity 50.0%; Pred. No. 9.75e+01;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hvdhkcsp 268  
 | :|:|  
 QY 481 HAIHRCHP 488

RESULT 8

ID R97209 standard; Protein; 4655 AA.  
 AC R97209;  
 DT 31-JUL-1996 (first entry)  
 DE Human placental calcium sensor protein.  
 KW Calcium sensor protein; CSP; hyperthyroidism; osteoporosis;  
 KW membranous glomerulonephritis; therapy; ulcer; wound; placenta.  
 OS Homo sapiens.  
 PN WO9615801-A1.  
 PD 30-MAY-1996.  
 PF 22-NOV-1995; U15203.  
 PR 23-NOV-1994; US-344836.  
 PR 07-JUN-1995; US-487314.  
 PA (RHON ) RHONE-POULENC RORER PHARM INC.

PI Amerstrom G, Crumley GR, Hjaln G, Juhlin C, Morse CC;  
 PI Murray EM, Rask L;  
 DR WPI: 96-268334/27.  
 DR N-PSDB: T28774.  
 PT Isolated human calcium sensor protein - useful develop prods. for  
 PT diagnosis and treatment of conditions, such as membranous  
 PT glomerulonephritis  
 PS Claim 7; Page 110-128; 236pp; English.  
 CC 500 kDa human calcium sensor proteins (CSP) (R97209-11) were  
 CC identified as the products of cDNA clones (T28774-76) respectively  
 CC obtd. from human placental, kidney and parathyroid cDNA libraries.  
 CC The CSPs show homology to LDL receptor proteins. CSP can be isolated  
 CC from parathyroid, placental or kidney tubule cells or obtd. by  
 CC expression of the encoding sequences in transformed host cells. It is  
 CC useful for identifying potential (ant)agonists and in the diagnosis  
 CC and treatment of e.g. membranous glomerulonephritis, and also in the  
 CC treatment of wounds and ulcers.  
 CC Sequence 4655 AA;  
 SQ

Query Match 67.6%; Score 46; DB 17; Length 4655;  
 Best Local Similarity 50.0%; Pred. No. 9.75e+01;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hvdhkcsp 268  
 | :|:|  
 QY 481 HAIHRCHP 488

RESULT 9

ID W43314 standard; Protein; 4655 AA.  
 AC W43314;  
 DT 08-JUN-1998 (first entry)  
 DE Human parathyroid calcium sensor protein.  
 KW Calcium sensor; human; glycoprotein 330; gp330; parathyroid;  
 KW low density lipoprotein; LDL; membranous glomerulonephritis;  
 KW osteoporosis; hyperthyroidism; therapy.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /label= Sig\_peptide  
 FT 26..4422  
 FT /note= "extracellular domain"  
 FT Domain 4423..4445  
 FT /note= "transmembrane domain"  
 FT Domain 4446..4655  
 FT /note= "C-terminal domain"  
 FT Misc\_difference 675 /note= "encoded by GTY"  
 FT Misc\_difference 1018 /note= "encoded by ACM"  
 FT WO9744050-A1.  
 PN 27-NOV-1997.  
 PD 20-MAY-1997; U08682.  
 PF 23-MAY-1996; US-652877.  
 PR (RHON ) RHONE-POULENC RORER PHARM INC.  
 PI Amerstrom G, Crumley GR, Hjaln G, Juhlin C, Morse CC,  
 PI Murray EM, Rask L;  
 DR WPI: 98-018219/02.  
 DR N-PSDB: V05997.  
 PT New human calcium sensor protein and vector encoding inhibitor of gp  
 PT 330 mediated transport of amyloid beta protein - useful to treat  
 PT membranous glomerulonephritis and central nervous system disorders  
 PS Claim 7; Page 177-195; 240pp; English.  
 CC This sequence comprises a human parathyroid calcium sensor protein  
 CC (hCSP), deduced from the nucleotide sequence (see V05996) of a  
 CC clone obtained by PCR amplification. Full-length sequences for  
 CC human placenta (see W43312) and kidney (see W43313) are also  
 CC claimed. hCSP is considered to be the human homologue of rat  
 CC Heymann nephritis antigen gp330 on the basis of sequence  
 CC similarity. It is proposed that the calcium sensor is a  
 CC universal sensor for recognition of variation in extracellular  
 CC calcium, and plays a key role in calcium regulation via different  
 CC organ systems. hCSP belongs to the low density lipoprotein

CC superfamily of glycoproteins. It can be expressed in a claimed  
 CC host cell using a vector comprising hCSP cDNA. Also claimed are a  
 CC method of using hCSP to treat human membranous glomerulonephritis,  
 CC and the use of hCSP to screen for agonists and antagonists of hCSP.  
 CC Such (ant)agonists are useful in the treatment of conditions  
 CC associated with perturbations in parathyroid hormone levels,  
 CC vitamin D3 production, oestrogen levels, osteoclast or osteoblast  
 CC activity, calcium secretion and calcium ion homeostasis, e.g.  
 CC hyperthyroidism and osteoporosis.  
 SQ Sequence 4655 AA;

Query Match 67.6%; Score 46; DB 29; Length 4655;

Best Local Similarity 50.0%; Pred. No. 9.75e+01;

Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 hvdhkcsp 268

QY 481 HAIRCHP 488

# RESULT 10

ID W8724 standard; Protein; 69 AA.

AC W8724;

DT 01-MAR-1999 (first entry)

DE Secreted protein encoded by gene 191 clone HJAB265.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN W09854963-A2.

PD 10-DEC-1998.

PF 04-JUN-1998; U11422.

PR 18-DEC-1997; US-070923.

PR 06-JUN-1997; US-048877.

PR 06-JUN-1997; US-048881.

PR 06-JUN-1997; US-048884.

PR 06-JUN-1997; US-048893.

PR 06-JUN-1997; US-048896.

PR 06-JUN-1997; US-048899.

PR 06-JUN-1997; US-048915.

PR 06-JUN-1997; US-048949.

PR 06-JUN-1997; US-048964.

PR 06-JUN-1997; US-048972.

PR 06-JUN-1997; US-049020.

PR 06-JUN-1997; US-049375.

PR 05-SEP-1997; US-057628.

PR 05-SEP-1997; US-057635.

PR 05-SEP-1997; US-057644.

PR 05-SEP-1997; US-057647.

PR 05-SEP-1997; US-057650.

PR 05-SEP-1997; US-057661.

PR 05-SEP-1997; US-057667.

PR 05-SEP-1997; US-057761.

PR 05-SEP-1997; US-057764.

PR 05-SEP-1997; US-057770.

PR 05-SEP-1997; US-057775.

PR 05-SEP-1997; US-057778.

PR 06-JUN-1997; US-048875.

PR 06-JUN-1997; US-048878.

PR 06-JUN-1997; US-048882.

PR 06-JUN-1997; US-048885.

PR 06-JUN-1997; US-048894.

PR 06-JUN-1997; US-048897.

PR 06-JUN-1997; US-048900.

PR 06-JUN-1997; US-048916.

PR 06-JUN-1997; US-048962.

PR 06-JUN-1997; US-048970.

PR 06-JUN-1997; US-048974.

PR 06-JUN-1997; US-049373.  
 PR 05-SEP-1997; US-057584.  
 PR 05-SEP-1997; US-057629.  
 PR 05-SEP-1997; US-057642.  
 PR 05-SEP-1997; US-057645.  
 PR 05-SEP-1997; US-057648.  
 PR 05-SEP-1997; US-057651.  
 PR 05-SEP-1997; US-057662.  
 PR 05-SEP-1997; US-057668.  
 PR 05-SEP-1997; US-057762.  
 PR 05-SEP-1997; US-057765.  
 PR 05-SEP-1997; US-057771.  
 PR 05-SEP-1997; US-057776.  
 PR 06-JUN-1997; US-048876.  
 PR 06-JUN-1997; US-048880.  
 PR 06-JUN-1997; US-048883.  
 PR 06-JUN-1997; US-048892.  
 PR 06-JUN-1997; US-048895.  
 PR 06-JUN-1997; US-048898.  
 PR 06-JUN-1997; US-048901.  
 PR 06-JUN-1997; US-048917.  
 PR 06-JUN-1997; US-048963.  
 PR 06-JUN-1997; US-048971.  
 PR 06-JUN-1997; US-049019.  
 PR 06-JUN-1997; US-049374.  
 PR 05-SEP-1997; US-057627.  
 PR 05-SEP-1997; US-057634.  
 PR 05-SEP-1997; US-057643.  
 PR 05-SEP-1997; US-057646.  
 PR 05-SEP-1997; US-057649.  
 PR 05-SEP-1997; US-057654.  
 PR 05-SEP-1997; US-057666.  
 PR 05-SEP-1997; US-057760.  
 PR 05-SEP-1997; US-057763.  
 PR 05-SEP-1997; US-057769.  
 PR 05-SEP-1997; US-057774.  
 PR 05-SEP-1997; US-057777.

(HUMA-) HUMAN GENOME SCI INC.

Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,

Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,

Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW, PI

Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,

Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;

WPI; 99-059865/05.

N-PSDB; V84601.

New isolated human genes and the secreted polypeptides they encode -  
 useful for diagnosis and treatment of e.g. cancers, neurological  
 disorders, immune diseases, inflammation or blood disorders  
 Claim 11; Page 578; 772pp; English.

The invention relates to nucleic acid sequences (V84411 to V84633)

encoding human secreted proteins (W88534 to W88756). The secreted protein  
 gene sequences are deposited with the ATCC under deposit numbers ATCC  
 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 cells comprising recombinant vectors containing the nucleic acid

sequences are used for the recombinant production of the secreted  
 proteins. The polynucleotide and amino acid sequences are useful for are  
 useful for preventing, treating or ameliorating medical conditions e.g.  
 by protein or gene therapy. Pathological conditions can be also diagnosed  
 by determining the amount of the new polypeptides in a sample or by  
 determining the presence of mutations in the new polynucleotides.

Specific uses are described for each of the polynucleotides, based on  
 which tissues they are most highly expressed in, and include developing  
 products for the diagnosis or treatment of cancer, neurodegenerative  
 disorders, developmental abnormalities and foetal deficiencies, blood  
 disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
 restenosis, prostate diseases, obesity, disorders involving osteoclasts  
 such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
 or thymus, digestive/endocrine disorders, infections and AIDS. The  
 polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents human secreted protein (see descriptor

```
CC line for gene number and clone identification).
SQ Sequence 69 AA;

Query Match 66.2%; Score 45; DB 38; Length 69;
Best Local Similarity 57.1%; Pred. No. 1.26e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 34 rplrch 40
QY 481 HAIHRCH 487

RESULT 11
ID W74745 standard; Protein; 84 AA.
AC W74745;
DT 25-JAN-1999 (first entry)
DE Human secreted protein encoded by gene 15 clone HE6DB26.
KW Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KW diagnosis; neurodegenerative disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 84 /label= unknown
FT WO9839448-A2.
PD 11-SEP-1998. U04493.
PF 06-MAR-1998; US-061060.
PR 02-OCT-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040326.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043576.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.

PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051926.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057669.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PR (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI: 98-506364/43.
DR N-PSDB; V59525.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 551; 721pp; English.
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 15 from the human cDNA clone HE6DB26
CC (deposited as clone ATCC 97897 and ATCC 209043).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
```

CC e.g. by protein or gene therapy. Also...pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 186 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see V59511 for described  
 CC uses).  
 SQ Sequence 84 AA;

Query Match 66.2%; Score 45; DB 37; Length 84;  
 Best Local Similarity 57.1%; Pred. No. 1.26e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 32 hvthrc 38  
 QY 481 HAIHRCH 487

## RESULT 12

ID R47861 standard; protein; 4544 AA.  
 AC R47861;  
 DT 20-JUL-1994 (first entry)  
 DE Alpha 2-Macroglobulin/LDL-receptor related protein.  
 KW Alpha 2-macroglobulin; Low Density Lipoprotein; LDL; receptor family;  
 KW LDL receptor related protein; LRP; small rhinovirus receptor; deriv;  
 KW minor rhinovirus; alpha2MR/LRP.  
 OS Homo sapiens.

FH Key Location/Qualifiers

FT misc\_difference 211..260 /note= "50 residues not shown in SEQ.ID.No.4"

FT misc\_difference 1990 /note= "Residue not shown in SEQ.ID.No.4"

FT misc\_difference 3050 /note= "Residue not shown in SEQ.ID.No.4"

FT WO9401553-A.

PN 20-JAN-1994.

PF 05-JUL-1993; E01728.

PR 08-JUL-1992; DE-222385.

PR 22-AUG-1992; DE-227892.

PR 19-FEB-1993; DE-305063.

PA (BOEHR) BOEHRINGER INGELHEIM INT GMBH.

PI Blaas D, Gruenberger M, Hofer F, Huettinger M, Kerjaschki D;

PI Kowalski H, Kuechler E, Machat H;

DR WPI; 94-035060/04.

PT New peptide derivs. of receptor for rhinovirus - of the small

PT receptor gp., and derived DNA, transformed cells and antibodies,

PT used e.g. to treat or prevent rhinovirus infection

PS Claim 5; Fig 2; 76pp; German.

CC Functional derivatives of members of the Minor Rhinovirus Receptor

CC group are claimed. The alpha-2 Macroglobulin/LDL-receptor related

CC protein of sequence R47861 (Herz et al.; EMBO J. 7;4119-4127 (1988))

CC is a preferred parent receptor. The derivs, which are preferably

CC soluble, extracellular forms of the native receptors, are useful

CC for treating and preventing viral (esp. rhinoviral) infections.

CC N.B. the SEQ.ID. listing includes a sequence (no.4) which differs

CC from the alpha2-MR/LRP sequence as indicated in the Features Table.

SQ Sequence 4544 AA;

Query Match 64.7%; Score 44; DB 9; Length 4544;

Best Local Similarity 57.1%; Pred. No. 1.63e+02;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1764 htinrc 1770

QY 481 HAIHRCH 487

## RESULT 13

ID R60517 standard; Protein; 4544 AA.

AC R60517;

DT 22-MAR-1995 (first entry)

DE Human alpha-2-MR.

KW Serine protease; Factor-Xa; recognition site;

KW fusion protein cleavage; protein folding; alpha-2-MR;

KW alpha-2-macroglobulin receptor/low density lipoprotein receptor.  
 OS Homo sapiens.

PN WO9418227-A.

PD 18-AUG-1994.

PF 04-FEB-1994; DK0054.

PR 04-FEB-1993; DK-000130.

PR 05-FEB-1993; DK-000139.

PR 03-DEC-1993; WO-G02492.

PA (DENZ-) DENZYME APS.

PI Etzerodt M, Holtet TL, Thogersen HC;

DR WPI; 94-279581/34.

PT Refolding of polypeptide molecules - using a cyclic process

PT involving denaturing and renaturing conditions to produce a

PT correctly folded prod

PS Disclosure; Page 131-146; 202pp; English.

CC Various domains and domain clusters of human alpha-2-MR protein

CC have been PCR amplified using the primers given in Q71252-65.

SQ Sequence 4544 AA;

Query Match 64.7%; Score 44; DB 11; Length 4544;

Best Local Similarity 57.1%; Pred. No. 1.63e+02;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1764 htinrc 1770

QY 481 HAIHRCH 487

## RESULT 14

ID W10404 standard; peptide; 9 AA.

AC W10404;

DT 21-AUG-1997 (first entry)

DE Antiphospholipid analogue peptide ACA-6501 phage library clone 2G12.

KW aPL; antibody mediated pathology; systemic lupus erythematosus;

KW antiphospholipid antibody syndrome; stroke; foetal loss; thrombosis.

OS Synthetic.

PN WO9640197-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09976.

PR 07-JUN-1995; US-482651.

PA (LJOL-) LA JOLLA PHARM CO.

PI Jones DS, Marquis DM, Victoria EJ, Yu L;

DR WPI; 97-051886/05.

PT New anti-phospholipid analogues - used to develop prods. for the

PT diagnosis and treatment of aPL antibody-mediated pathologies such as

PT systemic lupus erythematosus

PS Disclosure; Page 27; 118pp; English.

CC The present sequence is an antiphospholipid (aPL) analogue clone

CC 2G12 screened from ACA-6501 phage library. Peptides are derived

CC from screening random peptide phage libraries to identify

CC the sequences which best mimic the epitopes recognised by aPL

CC antibodies. A method has been produced for identifying analogues of

CC key epitopes recognised by aPL antibodies in patients suffering from

CC SPS, APS and other aPL antibody-mediated diseases. The products and

CC methods can be used for the diagnosis and treatment of aPL antibody-

CC mediated pathologies such as systemic lupus erythematosus,

CC antiphospholipid antibody syndrome, recurrent stroke, recurrent foetal

CC loss or thrombosis.

SQ Sequence 9 AA;

Query Match 63.2%; Score 43; DB 22; Length 9;

Best Local Similarity 83.3%; Pred. No. 2.11e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 4 haahrc 9

QY 481 HAIHRCH 486

## RESULT 15

ID W67128 standard; peptide; 9 AA.

AC W67128;

DT 18-DEC-1998 (first entry)

DE Antibody ACA-6501 epitope isolated from a phage library screen.  
KW antiphospholipid; aPL; analogue; stroke; foetal loss; B cell; epitope;  
KW thrombosis; antibody ACA-6501; conjugate; phage library screen;  
KW biopanning; micropanning.  
OS Synthetic.  
PN WO9746251-A1.  
PD 11-DEC-1997.  
PF 06-JUN-1997; U10075.  
PR 05-DEC-1996; US-760508.  
PR 06-JUN-1996; US-660092.  
PA (LJOL-) LA JOLLA PHARM CO.  
PI Jones DS, Marquis DM, Victoria EJ, Yu L;  
WPI; 98-051902/05.  
PT New anti-phospholipid analogues and conjugates containing them - are  
PT useful in treatment of disorders mediated by antiphospholipid  
PT antibody, such as stroke or foetal loss  
PS Disclosure: Page 26: 155pp; English.  
CC The invention relates to antiphospholipid (aPL) analogues that bind  
CC specifically to B cells to which an aPL epitope binds. Optimised  
CC analogues lack T cell epitope(s) and are useful as conjugates for  
CC treating aPL antibody-mediated diseases. Conjugates comprising aPL  
CC analogues and non-immunogenic valency platform molecules are provided as  
CC are novel non-immunogenic valency platform molecules and linkers. Methods  
CC of preparing and identifying the analogues, methods of treatment and  
CC methods and compositions for preparing conjugates of the analogues and  
CC diagnostic immunoassays for aPL antibodies are disclosed. The conjugates  
CC are useful for treatment of aPL antibody-mediated diseases, such as  
CC stroke, foetal loss, antiphospholipid antibody syndrome or thrombosis.  
CC Antibody ACA-6501, from a patient with a history of recurrent stroke,  
CC foetal loss, lupus and aortic valve replacements, was used in phage  
CC library screens. Sequences W67097-W67132 represent epitopes that  
CC were obtained in phage that micropanned.  
SQ Sequence 9 AA;

Query Match 63.2%; Score 43; DB 36; Length 9;  
Best Local Similarity 83.3%; Pred. No. 2.11e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 4 haahrc 9  
|||  
Qy 481 HAIHRC 486

Search completed: Thu May 20 13:17:04 1999  
Job time : 110 secs.

\*\*\*\*\*

WIREH

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 13:17:22 1999; MasPar time 3.83 Seconds

Tabular output not generated.

Title: >US-09-099-053-2

Description: (481-488) from US09099053.pep (6 of 6)

Perfect Score: 68

Sequence: 1 HAIHRCHP 8

Scoring table: PAM 150

Gap 11

Searched: 116695 seqs, 37453910 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.442; Variance 28.322; scale 0.722

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	50	73.5	594	2	S36502	2.27e+00
2	49	72.1	142	2	S21833	3.58e+00
3	49	72.1	321	1	S28390	3.58e+00
4	49	72.1	597	2	JC5829	3.58e+00
5	49	72.1	3412	1	GNWV7B	3.58e+00
6	49	72.1	3414	1	GNWVNE	3.58e+00
7	48	70.6	423	1	EHMSS	5.62e+00
8	48	70.6	731	2	B70770	5.62e+00
9	48	70.6	3415	2	A46105	5.62e+00
10	47	69.1	998	2	S78471	8.76e+00
11	47	69.1	998	2	S41397	8.76e+00
12	46	67.6	87	2	S38929	1.36e+01
13	46	67.6	197	2	B48290	1.36e+01
14	46	67.6	519	2	A49367	1.36e+01
15	46	67.6	581	2	JQ1879	1.36e+01
16	45	66.2	97	2	A28918	2.10e+01
17	45	66.2	283	3	JE0235	2.10e+01
18	45	66.2	345	2	S55969	2.10e+01
19	45	66.2	485	2	B40352	2.10e+01
20	45	66.2	573	2	E40247	2.10e+01
21	45	66.2	1056	1	WAD12	2.10e+01
22	45	66.2	1056	1	DJAP51	2.10e+01
23	44	64.7	369	2	S72734	3.21e+01

24	44	64.7	656	2	A41870	dnaA protein - Strept	3.21e+01
25	44	64.7	751	2	I48748	semaphorin E - mouse	3.21e+01
26	44	64.7	796	2	G64570	biotin sulfoxide redu	3.21e+01
27	44	64.7	1461	2	B70588	probable mbtf protein	3.21e+01
28	44	64.7	2638	1	A42545	genome polyprotein -	3.21e+01
29	44	64.7	2796	2	JC4743	fatty-acid synthase (	3.21e+01
30	44	64.7	4544	1	S02392	alpha-2-macroglobulin	3.21e+01
31	44	64.7	4545	1	S25111	alpha-2-macroglobulin	3.21e+01
32	43	63.2	440	1	TWBYM1	transcription factor	4.89e+01
33	43	63.2	454	2	A69017	dihydroorotase - Meth	4.89e+01
34	43	63.2	510	2	A42750	insulinoma-associated	4.89e+01
35	43	63.2	533	2	S42404	thromboxane-A synthas	4.89e+01
36	43	63.2	533	2	JN0683	thromboxane-A synthas	4.89e+01
37	43	63.2	880	2	S68828	finger protein XYL056	4.89e+01
38	43	63.2	1451	2	S65571	pattern formation pro	4.89e+01
39	43	63.2	1538	2	E70874	probable ppsB protein	4.89e+01
40	43	63.2	1548	2	S34583	serine proteinase (EC	4.89e+01
41	43	63.2	1765	2	A31494	DNA-directed RNA poly	4.89e+01
42	43	63.2	1765	2	B31494	DNA-directed RNA poly	4.89e+01
43	43	63.2	1766	2	B31875	DNA-directed RNA poly	4.89e+01
44	43	63.2	1766	2	A31875	DNA-directed RNA poly	4.89e+01
45	43	63.2	4572	2	S57908	hypothetical polyprot	4.89e+01

ALIGNMENTS

RESULT 1  
ENTRY S36502 #type complete  
TITLE L1 protein (alternative) - human papillomavirus type 27  
ORGANISM #formal\_name human papillomavirus type 27  
DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Sep-1997  
ACCESSIONS S36502  
REFERENCE S36469  
#authors Delius, H.; Hofmann, B.  
#submission submitted to the EMBL Data Library, August 1993  
#description Primer-directed sequencing of human papillomavirus types.  
#accession S36502  
#molecule\_type DNA  
#residues 1-594 #label DEL  
#cross-references EMBL:X7473; NID:g9396964; PID:g9396971  
CLASSIFICATION #superfamily papillomavirus L1 protein  
KEYWORDS late protein  
SUMMARY #length 594 #molecular-weight 67098 #checksum 7802

Query Match 73.5%; Score 50; DB 2; Length 594;  
Best Local Similarity 83.3%; Pred. No. 2.27e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 74 LHRCHP 79

Qy 483 IHRCHP 488

RESULT 2

ENTRY S21833 #type complete  
TITLE hypothetical protein - Mycobacterium tuberculosis  
ORGANISM #formal\_name Mycobacterium tuberculosis  
DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Feb-1995

ACCESSIONS S21833

REFERENCE S21832

#authors Patki, A.H.; Dale, J.W.  
#submission submitted to the EMBL Data Library, April 1991  
#accession S21833  
#status preliminary

#molecule\_type DNA

#residues 1-142 #label PAT

#cross-references EMBL:X59273

SUMMARY #length 142 #molecular-weight 15267 #checksum 3086

Query Match 72.1%; Score 49; DB 2; Length 142;  
Best Local Similarity 71.4%; Pred. No. 3.58e+00;

Matches	5;	Conservative	2;	Mismatches	0;	Indels	0;	Gaps	0;
Db	104	ALORCHP	110						
QY	482	AIHRCHP	488						
RESULT	3								
ENTRY		S28390	#type complete						
TITLE		homeotic protein mec-3 - <i>Caenorhabditis elegans</i>							
ORGANISM		#formal_name <i>Caenorhabditis elegans</i>							
DATE		17-Apr-1993	#sequence_revision 17-Apr-1993	#text_change					
ACCESSIONS		S28390; A27662							
REFERENCE		S28390							
#authors		Xue, D.; Finney, M.; Ruvkun, G.; Chalfie, M.							
#journal		EMBO J. (1992) 11:4969-4979							
#title		Regulation of the mec-3 gene by the C. elegans homeoproteins UNC-86 and MEC-3.							
#accession		S28390							
#status		nucleic acid sequence not shown							
#molecule_type		mRNA							
#residues		1-321	#label XUE						
#cross-references		EMBL:L02877; NID:gl56488; PID:gl56489							
REFERENCE		A27662							
#authors		Way, J.C.; Chalfie, M.							
#journal		Cell (1988) 54:5-16							
#title		mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans.							
#cross-references		MUID:88253425							
#accession		A27662							
#molecule_type		DNA							
#residues		'NPRLHDWLLT', 20, 'DLQESS', 28, 'ITASSKNSFTIIYFQ', 44-321							
#cross-references		GB:M20244; NID:gl56363; PID:gl56364							
GENETICS									
#gene		mec-3							
#introns		82/3; 106/3; 190/3; 228/3; 300/1							
CLASSIFICATION		#superfamily homeotic protein mec-3; homeobox homology; LIM metal-binding repeat homology							
KEYWORDS		DNA binding; duplication; homeobox; nucleus; transcription regulation; zinc							
FEATURE									
29-79		#domain LIM metal-binding repeat homology #label LIM1\							
89-145		#domain LIM metal-binding repeat homology #label LIM2\							
218-274		#domain homeobox homology #label HOX							
SUMMARY		#length 321 #molecular-weight 37088 #checksum 1652							
Query Match		72.1%; Score=49.7; DB:1; - Length 321;							
Best Local Similarity		83.3%; Pred. No. 3.58e+00;							
Matches	5;	Conservative	1;	Mismatches	0;	Indels	0;	Gaps	0;
Db	84	HSIIRC	89						
QY	481	HAIRHC	486						
RESULT	4								
ENTRY		JC5829	#type complete						
TITLE		alkylglycerone-phosphate synthase (EC 2.5.1.26) - <i>Caenorhabditis elegans</i>							
ALTERNATE_NAMES		alkyldihydroxyacetonephosphate synthase							
ORGANISM		#formal_name <i>Caenorhabditis elegans</i>							
DATE		20-Feb-1998	#sequence_revision 13-Mar-1998	#text_change					
ACCESSIONS		JC5829							
REFERENCE		JC5829							
#authors		de Vet, E.C.J.M.; Prinsen, H.C.M.T.; van den Bosch, H.							
#journal		Biochem. Biophys. Res. Commun. (1998) 242:277-281							
#title		Nucleotide sequence of a cDNA clone encoding a <i>Caenorhabditis elegans</i> homolog of mammalian alkyldihydroxyacetonephosphate synthase; Evolutionary							

```

144,434,641,753,
861,983,999,1228,
1649,1988,2044,
2052,2447,2466,
2685,2725
#label NS5\
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 3412 #molecular-weight 377979 #checksum 7007
Query Match 72.1%; Score 49; DB 1; Length 3412;
Best Local Similarity 50.0%; Pred. No. 3.58e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HEVHOCOP 1604
QY 481 HAIHRCHP 488

RESULT 6
ENTRY GNMVNE #type complete
TITLE genome polyprotein - tick-borne encephalitis virus (subtype
Western, strain Neudoerfl)
CONTAINS capsid protein C; envelope protein E; membrane protein M;
nonstructural protein NS1; nonstructural protein NS2a;
nonstructural protein NS2b; nonstructural protein NS3;
nonstructural protein NS4a; nonstructural protein NS4b;
nonstructural protein NS5
ORGANISM #formal_name tick-borne encephalitis virus
DATE 31-Dec-1989 #sequence_revision 30-Jun-1991 #text_change
15-May-1998
ACCESSIONS A31052; A32596
REFERENCE Mandi, C.W.; Heinz, F.X.; Kunz, C.
Virology (1988) 166:197-205
#journal Sequence of the structural proteins of tick-borne
#title encephalitis virus (Western subtype) and comparative
analysis with other flaviviruses.
#cross-references MUID:88322870
#accession A31052
#molecule_type genomic RNA
#residues 1-779 #label MAN1
#cross-references GB:M21498
REFERENCE A32596
#authors Mandi, C.W.; Heinz, F.X.; Stoeckl, E.; Kunz, C.
#journal Virology (1989) 173:291-301
#title Genome sequence of tick-borne encephalitis virus (Western
subtype) and comparative analysis of nonstructural proteins
with other flaviviruses.
#cross-references MUID:90051080
#accession A32596
#molecule_type genomic RNA
#residues 767-3414 #label MAN2
#cross-references GB:M27157
CLASSIFICATION #superfamily yellow fever virus genome polyprotein; DEAD/H
box helicase homology
KEYWORDS ATP; capsid protein; envelope protein; glycoprotein;
nonstructural protein; P-loop; polyprotein; transmembrane
protein
FEATURE
2-116 #product capsid protein C #status predicted #label CPC\
117-280 #product membrane protein M precursor #status predicted
#label MPP\
117-205 #domain nonterminal signal sequence #status predicted
#label SIG\
206-280 #product membrane protein M #status predicted #label
MPM\
246-264 #domain transmembrane #status predicted #label TM1\
281-776 #product envelope protein E #status predicted #label
EPE\
738-751 #domain transmembrane #status predicted #label TM2\
777-1128 #product nonstructural protein NS1 #status predicted
#label NS1\
1129-1358 #product nonstructural protein NS2a #status predicted
#label NS2A\
#product nonstructural protein NS2b #status predicted
#label NS2B\
#product nonstructural protein NS3 #status predicted
#label NS3\
#domain DEAD/H box helicase homology #label DEAD\
#region nucleotide-binding motif A (P-loop)\
#region nucleotide-binding motif B\
#region DEAH motif\
#product nonstructural protein NS4a #status predicted
#label NS4A\
#product nonstructural protein NS4b #status predicted
#label NS4B\
#product nonstructural protein NS5 #status predicted
#label NS5\
144,434,641,753,
861,983,999,1649,
1988,2044,2447,
2529,2686,2726
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 3414 #molecular-weight 378383 #checksum 1083
Query Match 72.1%; Score 49; DB 1; Length 3414;
Best Local Similarity 50.0%; Pred. No. 3.58e+00;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HEVHOCOP 1604
QY 481 HAIHRCHP 488

RESULT 7
ENTRY EHMSS #type complete
TITLE Ig epsilon chain C region (version 2) - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change
16-Aug-1996
ACCESSIONS A02145
REFERENCE Ishida, N.; Ueda, S.; Hayashida, H.; Miyata, T.; Honjo, T.
EMBO J. (1982) 1:1117-1123
#journal The nucleotide sequence of the mouse immunoglobulin epsilon
#title gene: comparison with the human epsilon gene sequence.
#cross-references MUID:84236092
#accession A02145
#molecule_type DNA
#residues 1-423 #label ISH
#note the sequence was determined from the germline gene
GENETICS
#introns 91/1; 199/1; 307/1
COMPLEX An immunoglobulin heterotetramer subunit consists of two
identical light (kappa or lambda) and two identical heavy
(alpha, delta, epsilon, gamma, or mu) chains usually
stabilized by interchain disulfide bonds. In some cases,
such as IgA and IgM, the subunits associate into larger
oligomers.
CLASSIFICATION #superfamily immunoglobulin C region; immunoglobulin homology
KEYWORDS duplication; glycoprotein; heterotetramer; immunoglobulin
FEATURE
16-77 #domain immunoglobulin homology #label IM1\
115-183 #domain immunoglobulin homology #label IM2\
220-288 #domain immunoglobulin homology #label IM3\
325-396 #domain immunoglobulin homology #label IM4\
223-75,122-181,
227-286,332-394
43,84,167,239,262,
#disulfide_bonds #status predicted\
417 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 423 #molecular-weight 47681 #checksum 7738
Query Match 70.6%; Score 48; DB 1; Length 423;
Best Local Similarity 71.4%; Pred. No. 5.62e+00;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```



```
Db 105 SIHRCOP 111
QY 482 AIHRCHP 488

RESULT 8
ENTRY #type complete
TITLE probable glgB protein - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS B70770
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MUID:98295987
#accession B70770
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-731 #label COL
#cross-references GB:273902; GB:AL123456; NID:g3261576; PID:e245020; PID:g1340092
#experimental_source strain H37Rv

GENETICS
#gene glgB
#summary #length 731 #molecular-weight 81729 #checksum 2172

Query Match 70.6%; Score 48; DB 2; Length 731;
Best Local Similarity 62.5%; Pred. No. 5.62e+00;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 612 NDIYRCHP 619
QY 481 HAIHRCHP 488

RESULT 9
ENTRY #type complete
TITLE polyprotein(NS1, NS3, NS5, NS2A, NS2B, NS4A, NS4B, small capsid protein C, large envelope protein E, membrane-anchored protein M precursor) - Powassan virus
ORGANISM #formal_name Powassan virus
DATE 07-Apr-1994 #sequence_revision 25-Apr-1997 #text_change
ACCESSIONS A46105
REFERENCE A46105
#authors Mandl, C.W.; Holzmann, H.; Kunz, C.; Heinz, F.X.
#journal Virology (1993) 194:173-184
#title Complete genomic sequence of Powassan virus: evaluation of genetic elements in tick-borne versus mosquito-borne flaviviruses.
#cross-references MUID:93242744
#accession A46105
#status preliminary
#molecule_type genomic RNA
#residues 1-3415 #label MAN
#cross-references GB:I06436; NID:g309916; PID:g309917
#note sequence extracted from NCBI backbone (NCBIN:130654, NCBIP:130655)

CLASSIFICATION
#superfamily yellow fever virus genome polyprotein; DEAD/H box helicase homology
ATP; P-loop
KEYWORDS
FEATURE 1688-1956 #domain DEAD/H box helicase homology #label DEAD\
1688-1695 #region nucleotide-binding motif A (P-loop)\
1776-1781 #region nucleotide-binding motif B\
1780-1783 #region DEAD motif
SUMMARY #length 3415 #molecular-weight 378568 #checksum 9967

Query Match 70.6%; Score 48; DB 2; Length 3415;
Best Local Similarity 62.5%; Pred. No. 5.62e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HKIHQCOP 1604
QY 481 HAIHRCHP 488

RESULT 10
ENTRY #type complete
TITLE hypothetical protein A - black beetle virus
ORGANISM #formal_name black beetle virus
DATE 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change
ACCESSIONS S78471; S28728
REFERENCE S78471
#authors Dasgupta, R.
#submission submitted to the EMBL Data Library, July 1986
#accession S78471
#molecule_type genomic RNA
#residues 1-998 #label DAS
#cross-references EMBL:X02396; NID:g60679; PID:g60680
REFERENCE S28728
#authors Dasgupta, R.; Dasgupta, R.; Ghosh, A.; Kaesberg, P.
#journal J. Mol. Biol. (1985) 182:183-189
#title Structure of the black beetle virus genome and its functional implications.
#accession S28728
#molecule_type genomic RNA
#residues 1-883 'AALRTPTWTRYQC' #label DAW
#cross-references EMBL:X02396
KEYWORDS RNA biosynthesis
SUMMARY #length 998 #molecular-weight 112196 #checksum 4131

Query Match 69.1%; Score 47; DB 2; Length 998;
Best Local Similarity 62.5%; Pred. No. 8.76e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 257 HKIHCRP 264
QY 481 HAIHRCHP 488

RESULT 11
ENTRY #type complete
TITLE protein A - flock house virus
ORGANISM #formal_name flock house virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
ACCESSIONS S41397
REFERENCE S41397
#authors Dasgupta, R.
#submission submitted to the EMBL Data Library, January 1994
#description Near identity in the polymerase gene of two serologically distinct nodaviruses.
#accession S41397
#status preliminary
#molecule_type genomic RNA
#residues 1-998 #label DAS
#cross-references EMBL:X77156; NID:g450500; PID:g450501
SUMMARY #length 998 #molecular-weight 112187 #checksum 4283
```

```
Query Match      69.1%; Score 47.1; DB 2; Length 998;
Best Local Similarity 62.5%; Pred. No. 8.76e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 257 HKTHCRP 264
| : : : :
QY 481 HAIHRCHP 488

RESULT 12
ENTRY S38929 #type fragment
TITLE pdxj protein - Aquifex pyrophilus (strain K01SA) (fragment)
ORGANISM #formal_name Aquifex pyrophilus
#variety strain K01SA
DATE 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change
06-Jun-1997
ACCESSIONS S38929
REFERENCE S38927
#authors Bocchetta, M.
#submission submitted to the EMBL Data Library, July 1993
#accession S38929
##molecule_type DNA
##residues 1-87 #label BOC
##cross-references EMBL:X74277
##experimental_source strain K01SA
GENETICS
#gene pdxj
#start_codon GTG
SUMMARY #length 87 #checksum 4629

Query Match      67.6%; Score 46; DB 2; Length 87;
Best Local Similarity 66.7%; Pred. No. 1.36e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 36 HPVHRC 41
| : : : :
QY 481 HAIHRC 486

RESULT 13
ENTRY B48290 #type complete
TITLE hypothetical protein 197 - Helicobacillus mobilis
ORGANISM #formal_name Helicobacillus mobilis
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
19-Dec-1997
ACCESSIONS B48290
REFERENCE B48290
#authors Liebl, U.; Mockensturm-Wilson, M.; Trost, J.T.; Brune, D.C.;
Blankenship, R.E.; Vermaas, W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7124-7128
#title Single core polypeptide in the reaction center of the
photosynthetic bacterium Helicobacillus mobilis: structural
implications and relations to other photosystems.
#accession B48290
##status preliminary; not compared with conceptual translation
##molecule_type DNA
##residues 1-197 #label LIE
##cross-references GB:L19604
SUMMARY #length 197 #molecular-weight 22200 #checksum 1191

Query Match      67.6%; Score 46; DB 2; Length 197;
Best Local Similarity 83.3%; Pred. No. 1.36e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 48 QATHRC 53
| : : : :
QY 481 HAIHRC 486

RESULT 14
ENTRY A49367 #type complete
TITLE transducin homolog sazd - human
ORGANISM #formal_name Homo sapiens #common_name man
```

```
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
09-Apr-1998
ACCESSIONS A49367
REFERENCE A49367
#authors Weinstat-Saslow, D.L.; Germino, G.G.; Somlo, S.; Readers,
S.T.
#journal Genomics (1993) 18:709-711
#title A transducin-like gene maps to the autosomal dominant
polycystic kidney disease gene region.
#accession A49367
##status preliminary
##molecule_type mRNA
##residues 1-519 #label WEI
##cross-references GB:U02609; NID:G414535; PID:G414536
GENETICS
#gene sazd
CLASSIFICATION #superfamily WD repeat homology
FEATURE
77-110 #domain WD repeat homology #label WD1\
258-291 #domain WD repeat homology #label WD2\
361-394 #domain WD repeat homology #label WD3
SUMMARY #length 519 #molecular-weight 56047 #checksum 1718

Query Match      67.6%; Score 46; DB 2; Length 519;
Best Local Similarity 57.1%; Pred. No. 1.36e+01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 358 HTQORCH 364
| : : : :
QY 481 HAIHRCH 487

RESULT 15
ENTRY JQ1879 #type complete
TITLE hypothetical 65.4K protein - pea enation mosaic virus
ORGANISM #formal_name pea enation mosaic virus, PEMV
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
12-May-1995
ACCESSIONS JQ1879
REFERENCE JQ1878
#authors Demler, S.A.; Rucker, D.G.; de Zoeten, G.A.
#journal J. Gen. Virol. (1993) 74:1-14
#title The chimeric nature of the genome of pea enation mosaic
virus: The independent replication of RNA2.
#accession JQ1879
##molecule_type mRNA
##residues 1-581 #label DEM
##experimental_source strain WSG
GENETICS
#map_position segment RNA2
#start_codon CTG
CLASSIFICATION #superfamily barley yellow dwarf virus RNA-directed RNA
polymerase homology
FEATURE
256-428 #domain barley yellow dwarf virus RNA-directed RNA
polymerase homology #label BYD
SUMMARY #length 581 #molecular-weight 65404 #checksum 4973

Query Match      67.6%; Score 46; DB 2; Length 581;
Best Local Similarity 57.1%; Pred. No. 1.36e+01;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 8 HVVHOCH 14
| : : : :
QY 481 HAIHRCH 487

Search completed: Thu May 20 13:18:43 1999
Job time : 81 secs.
```

\*\*\*\*\*  
WATERMAN  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 13:19:01 1999; MasPar time 2.42 Seconds  
Tabular output not generated. ....93.310 Million\*cell\*updates/sec

Title: >US-09-099-053-2  
Description: (481-488) from US09099053.pep (6 of 6)  
Perfect Score: 68  
Sequence: 1 HAHRCHP 8

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 21.282; Variance 25.495; scale 0.835

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result	No.	Score	Query Match Length DB ID	Description	
1	50	73.5	594	1 VLL_HPV27...	7.71e-01
2	49	72.1	321	1 MEC3_CAEEL MECHANOSENSORY PROTEIN	1.28e+00
3	49	72.1	3412	1 POLG_TBEVS GENOME POLYPROTEIN [CO	1.28e+00
4	49	72.1	3414	1 POLG_TBEVW GENOME POLYPROTEIN [CO	1.28e+00
5	49	72.1	3414	1 POLG_TBEVH GENOME POLYPROTEIN [CO	1.28e+00
6	48	70.6	376	1 LYMP_MYCTU LYMPHOPAIN PRECURSOR (	2.12e+00
7	48	70.6	731	1 GLGB_MYCTU PROBABLE 1,4-ALPHA-GLU	2.12e+00
8	48	70.6	3415	1 POLG_POWVL GENOME POLYPROTEIN [CO	2.12e+00
9	46	67.6	519	1 SAZD_HUMAN WD-REPEAT PROTEIN SAZD	5.66e+00
10	46	67.6	1551	1 VQ12_CAEEL HYPOTHETICAL 175.7 KD	5.66e+00
11	46	67.6	4655	1 LRP2_HUMAN LOW-DENSITY LIPOPROTEIN	5.66e+00
12	45	66.2	345	1 VATX_YEAST VACUOLAR ATP SYNTHASE	9.16e+00
13	45	66.2	485	1 BIND_STRFN BINDIN PRECURSOR	9.16e+00
14	45	66.2	1056	1 DPOL_ADE05 DNA POLYMERASE (EC 2.7	9.16e+00
15	45	66.2	1056	1 DPOL_ADE02 DNA POLYMERASE (EC 2.7	9.16e+00
16	44	64.7	350	1 LEU3_AZOVI 3-ISOPROPYLMALATE DEHY	1.47e+01
17	44	64.7	369	1 RUVE_MYCLE HOLLIDAY JUNCTION DNA	1.47e+01
18	44	64.7	656	1 DNAA_STRCO CHROMOSOMAL REPLICATIO	1.47e+01
19	44	64.7	3414	1 POLG_LANVT GENOME POLYPROTEIN [CO	1.47e+01
20	44	64.7	4544	1 LRPI_HUMAN LOW-DENSITY LIPOPROTEI	1.47e+01
21	43	63.2	320	1 IY13_NPVOP HYPOTHETICAL 36.6 KD P	2.35e+01
22	43	63.2	438	1 ILL1_ARATH IAA-AMINO ACID HYDROLA	2.35e+01
23	43	63.2	439	1 ILL2_ARATH IAA-AMINO ACID HYDROLA	2.35e+01

24	43	63.2	440	1 NAM1_YEAST NAM1 PROTEIN PRECURSOR	2.35e+01
25	43	63.2	454	1 PYRC_METH DIHYDROOROTASE (EC 3.5	2.35e+01
26	43	63.2	484	1 RGS9_BOVIN REGULATOR OF G-PROTEIN	2.35e+01
27	43	63.2	510	1 IAL1_HUMAN ZINC FINGER PROTEIN IA	2.35e+01
28	43	63.2	533	1 THAS_MOUSE THROMBOXANE-A SYNTHASE	2.35e+01
29	43	63.2	533	1 THAS_RAT THROMBOXANE-A SYNTHASE	2.35e+01
30	43	63.2	547	1 CO9_HORSE COMPLEMENT COMPONENT C	2.35e+01
31	43	63.2	595	1 P2X7_RAT P2X PURINOCEPTOR 7 (AT	2.35e+01
32	43	63.2	880	1 YJF6_YEAST HYPOTHETICAL 98.9 KD Z	2.35e+01
33	43	63.2	1043	1 CHS2_PARBH CHITIN SYNTHASE 2 (EC	2.35e+01
34	43	63.2	1451	1 EM30_ARATH PATTERN FORMATION PROT	2.35e+01
35	43	63.2	1766	1 RPB1_TRYBB DNA-DIRECTED RNA POLYM	2.35e+01
36	43	63.2	1766	1 RPB2_TRYBB DNA-DIRECTED RNA POLYM	2.35e+01
37	42	61.8	121	1 INL3_MOUSE LEYDIG INSULIN-LIKE PE	3.71e+01
38	42	61.8	344	1 YDTG_SCHPO HYPOTHETICAL 39.8 KD P	3.71e+01
39	42	61.8	372	1 VGLI_HSV23 GLYCOPROTEIN I.	3.71e+01
40	42	61.8	383	1 AMCL_ORYSA ALPHA-AMYLASE ISOZYME	3.71e+01
41	42	61.8	542	1 GLPA_ECOLI ANAEROBIC GLYCEROL-3-P	3.71e+01
42	42	61.8	682	1 DYNEIN_78 KD INTERMED	3.71e+01
43	42	61.8	1231	1 CFAH_HUMAN COMPLEMENT FACTOR H PR	3.71e+01
44	42	61.8	1564	1 PDRA_YEAST ATP-DEPENDENT PERMEASE	3.71e+01
45	42	61.8	4753	1 LRP_CAEEL LOW-DENSITY LIPOPROTEI	3.71e+01

ALIGNMENTS

RESULT 1  
ID VLL\_HPV27 STANDARD; PRT; 594 AA.  
AC P36736; O81959;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE MAJOR CAPSID PROTEIN L1.  
GN L1.  
OS HUMAN PAPILLOMAVIRUS TYPE 27.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94265501.  
RA DELIUS H., HOFMANN B.;  
RT "Primer-directed sequencing of human papillomavirus types."  
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).  
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-110 IS THE  
CC INITIATOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; X74473; G396971;  
CC EMBL; X74473; G396972; ALT\_INIT.  
CC PIR; S36502; S36502.  
CC PFAM; PF00500; late\_protein\_L1; 1.  
CC COAT\_PROTEIN; LATE PROTEIN.  
CC KW COAT PROTEIN; LATE PROTEIN.  
CC SQ SEQUENCE 594 AA; 67098 MW; 18757AA1 CRC32;  
CC -----

Query Match 73.5%; Score 50; DB 1; Length 594;  
Best Local Similarity 83.3%; Pred. No. 7.71e-01;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 74 LHRCHP 79  
:|||||  
QY 483 IHRCHP 488

RESULT 2  
ID MEC3\_CAEEL STANDARD; PRT; 321 AA.  
AC P09088;  
DT 01-NOV-1988 (REL. 09, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE MECHANOSENSORY PROTEIN 3.  
GN MEC-3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88253425.  
RA WAT J.C., CHALFIE M.;  
RT "mec-3, a homeobox-containing gene that specifies differentiation of  
RL the touch receptor neurons in C. elegans.";  
CELL 54:5-16(1988).  
RN [2]  
RP REVISIONS. SEQUENCE FROM N.A.  
RX MEDLINE; 93099872.  
RA XUE D., FINNEY M., RUVKUN G., CHALFIE M.;  
RT "Regulation of the mec-3 gene by the C.elegans homeoproteins UNC-86  
RL and MEC-3.";  
EMBO J. 11:4969-4979(1992).  
CC -!- FUNCTION: MEC-3 SPECIES DIFFERENTIATION OF THE SET OF SIX TOUCH  
CC RECEPTOR NEURONS. BINDS COOPERATIVELY AS A HETERODIMER WITH UNC-86  
CC TO SITES IN THE MEC-3 GENE PROMOTER.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: TO OTHER HOMEBOX DOMAINS. BELONGS TO THE LIM  
CC SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC  
CC IONS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L02877; G156489; -  
DR EMBL; M20244; G156384; ALT\_SEQ.  
DR PIR; A27662; A27662.  
DR PIR; S28390; S28390.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00478; LIM\_DOMAIN\_1; 2.  
DR PROSITE; PS00023; LIM\_DOMAIN\_2; 2.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
DR PFAM; PF00046; homeobox; 1.  
DR PFAM; PF00412; LIM; 2.  
DR HSSP; P06601; LFJL.  
DR TRANSFAC; T01076; -  
KW HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN;  
KW REPEAT; LIM MOTIF; METAL-BINDING; ZINC; -----  
FT DOMAIN 29 79  
FT DOMAIN 89 145  
FT DNA\_BIND 217 276  
FT DOMAIN 315 321 ASP/GLU-RICH (ACIDIC).  
SQ SEQUENCE 321 AA; 37088 MW; 24F3D4B9 CRC32;  
Query Match 72.1%; Score 49; DB 1; Length 321;  
Best Local Similarity 83.3%; Pred. No. 1.28e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 84 HSIHRC 89  
Qy 481 HAIHRC 486  
RESULT 3  
ID POLG-TBEVS STANDARD; PRT; 3412 AA.  
AC P07720; P07721; Q88475; Q88476; Q88477; Q88478; Q88479; Q88877;  
DT 01-APR-1988 (REL. 07, CREATED)  
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX  
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL  
DE PROTEINS NS1, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED  
DE RNA POLYMERASE (EC 2.7.7.48) (NS5)].  
OS TICK-BORNE ENCEPHALITIS VIRUS (STRAIN SORJIN) (TBEV).  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;  
OC FLAVIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90101381.  
RA PLETNEV A.G., YAMSHCHIKOV V.F., BLINOV V.M.;  
RT "Nucleotide sequence of the genome and complete amino acid sequence  
RL of the polyprotein of tick-borne encephalitis virus.";  
VIROLOGY 174:250-263(1990).  
RN [2]  
RP SEQUENCE OF 1-1190 FROM N.A.  
RX MEDLINE; 88319988.  
RA YAMSHCHIKOV V.F., PLETNEV A.G.;  
RT "Nucleotide sequence of the genome region encoding the structural  
RL proteins and the NS1 protein of the tick borne encephalitis virus.";  
NUCLEIC ACIDS RES. 16:7750-7750(1988).  
RN [3]  
RP SEQUENCE OF 1-683 AND 777-1002 FROM N.A.  
RX MEDLINE; 86220766.  
RA PLETNEV A.G., YAMSHCHIKOV V.F., BLINOV V.M.;  
RT "Tick-borne encephalitis virus genome. The nucleotide sequence coding  
RL for virion structural proteins.";  
FEBS LETT. 200:317-321(1986).  
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA.  
CC -!- THE NONSTRUCTURAL PROTEINS NS1 PRESENTS TWO ALTERNATIVE CLEAVAGE  
CC SITES FOR ITS C-TERMINUS, WHICH MAY DEFINE A SOLUBLE OR A  
CC MEMBRANE-BOUND FORM OF NS1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X07755; G62026; -  
DR EMBL; X07755; G62027; ALT\_TERM.  
DR EMBL; X07755; E11928; ALT\_SEQ.  
DR EMBL; X07755; E11929; ALT\_SEQ.  
DR EMBL; X07755; E11930; ALT\_SEQ.  
DR EMBL; X07755; E11931; ALT\_SEQ.  
DR EMBL; X03870; G62107; -  
DR EMBL; X03870; E12099; ALT\_SEQ.  
DR EMBL; X03870; E12100; ALT\_SEQ.  
DR EMBL; X03870; E12101; ALT\_SEQ.  
DR EMBL; X03871; E12103; -  
DR PIR; A33776; GNVVTB.  
DR PIR; A24055; A24055.  
DR PFAM; PF00869; Flavi\_glycoprot; 1.  
DR PFAM; PF00948; Flavi\_NS1; 1.  
DR PFAM; PF00949; Flavi\_helicase; 1.  
DR PFAM; PF00972; Flavi\_NS5; 1.  
DR PFAM; PF01002; Flavi\_NS2B; 1.  
DR PFAM; PF01003; Flavi\_capsid; 1.  
DR PFAM; PF01004; Flavi\_M; 1.  
DR PFAM; PF01005; Flavi\_NS2A; 1.  
DR HSSP; P14336; 1SVB.  
KW POLYPROTEIN; GLYCOPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE;  
KW CORE PROTEIN; COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING;  
KW TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.  
RN INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE





Db 612 NDIYRCHP 619  
: | | | |

ET	TRANSMEM	1294	13
ET	TRANSMEM	1295	14

FT TRANSMEM 1453 1473 POTENTIAL.  
FT TRANSMEM 2161 2184 POTENTIAL.  
FT TRANSMEM 2191 2211 POTENTIAL.  
FT TRANSMEM 2244 2260 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).  
FT TRANSMEM 2347 2367 POTENTIAL.  
FT TRANSMEM 2434 2454 POTENTIAL.  
FT NP\_BIND 1688 1695 ATP (BY SIMILARITY).  
FT SITE 1780 1783 DEAH BOX.  
FT DISULFID 281 308 BY SIMILARITY.  
FT DISULFID 338 394 BY SIMILARITY.  
FT DISULFID 352 383 BY SIMILARITY.  
FT DISULFID 370 399 BY SIMILARITY.  
FT DISULFID 464 568 BY SIMILARITY.  
FT DISULFID 585 617 BY SIMILARITY.  
FT CARBOHYD 142 142 POTENTIAL.  
FT CARBOHYD 432 432 POTENTIAL.  
FT CARBOHYD 860 860 POTENTIAL.  
FT CARBOHYD 983 983 POTENTIAL.  
FT CARBOHYD 999 999 POTENTIAL.  
SQ SEQUENCE 3415 AA; 378564 MW; A6887372 CRC32;

Query Match 70.6%; Score 48; DB 1; Length 3415;  
Best Local Similarity 62.5%; Pred. No. 2.12e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HKHQCP 1604  
| : : : :  
QY 481 HAIHRCHP 488

RESULT 9  
ID SAZD\_HUMAN STANDARD; PRT; 519 AA.  
AC Q12788;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DE WD-REPEAT PROTEIN SAZD.  
GN SAZD.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94140377.  
RA WEINSTAT-SASLOW D.L., GERMINO G.G., SOMLO S., REEDERS S.T.;  
RT "A transducin-like gene maps to the autosomal dominant polycystic  
kidney disease gene region.";  
RL GENOMICS 18:709-711(1993).  
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U02609; G414536; -  
DR PROSITE; PS00678; WD-REPEATS; 2.  
DR PFAM; PF00400; G-beta; 5.  
KW REPEAT; WD REPEAT.  
FT REPEAT 1 23 WD1.  
FT REPEAT 35 67 WD2.  
FT REPEAT 79 109 WD3.  
FT REPEAT 260 290 WD4.  
FT REPEAT 305 336 WD5.  
FT REPEAT 363 393 WD6.  
FT REPEAT 447 477 WD7.  
SQ SEQUENCE 519 AA; 56047 MW; 2C1429C5 CRC32;  
  
Query Match 67.6%; Score 46; DB 1; Length 519;  
Best Local Similarity 57.1%; Pred. No. 5.66e+00;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 358 HTTQCH 364  
| : : : :  
QY 481 HAIHRCH 487  
  
RESULT 10  
ID YQ12\_CAEEL STANDARD; PRT; 1551 AA.  
AC Q09449;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 175.7 KD PROTEIN C05C10.2 IN CHROMOSOME II.  
GN C05C10.2.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA MATTHEWS P.;  
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z48178; E1351581; -  
DR WORMPEP; C05C10.2; CE01466.  
KW HYPOTHETICAL PROTEIN; ATP-BINDING; HELICASE.  
FT NP\_BIND 1021 1028 ATP (BY SIMILARITY).  
SQ SEQUENCE 1551 AA; 175699 MW; 28399662 CRC32;  
  
Query Match 67.6%; Score 46; DB 1; Length 1551;  
Best Local Similarity 50.0%; Pred. No. 5.66e+00;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 1276 RNVRCHP 1283  
| : : : :  
QY 481 HAIHRCHP 488  
  
RESULT 11  
ID LRP2\_HUMAN STANDARD; PRT; 4655 AA.  
AC P98164; Q00711; Q16215;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 2 PRECURSOR (MEGALIN)  
(GLYCOPROTEIN 330).  
GN LRP2.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY.  
RX MEDLINE: 96305376.  
RA HJAEML G., MURRAY E., CRUMLEY G., HARAZIM W., LUNDGREN S., ONYANGO I.,  
RA EK B., LARSSON M., JUHLIN C., HELLMAN P., DAVIS H., AAKERSTROEM G.,  
RA RASK L., MORSE B.;  
RT "Cloning and sequencing of human gp330, a Ca(2+)-binding receptor  
with potential intracellular signaling properties.";  
RL EUR. J. BIOCHEM. 239:132-137(1996).  
RN [2]  
RP SEQUENCE OF 2705-4453 FROM N.A.  
RC TISSUE-KIDNEY;



RA KNAK C., ARGRAVES W.S.;  
RL SUBMITTED (DEC-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE OF 3833-4453 FROM N.A.  
RX TISSUE-KIDNEY;  
RC MEDLINE; 95048397.  
RA KORENBERG J.R., ARGRAVES K.M., CHEN X.N., TRAN H.,  
RA STRICKLAND D.K., ARGRAVES W.S.;  
RT "Chromosomal localization of human genes for the LDL receptor family  
RT member glycoprotein 330 (LRP2) and its associated protein RAP  
RT (LRPAP1).";  
RL GENOMICS 22:88-93(1994).  
RN [4]  
RP SEQUENCE OF 4139-4406 FROM N.A.  
RX MEDLINE; 94244704.  
RA LUNDGREN S., HJALM G., HELLMAN P., EK B., JUHLIN C., RASTAD J.,  
RA KLARESKOG L., AKERSTROM G., RASK L.;  
RT "A protein involved in calcium sensing of the human parathyroid and  
RT placental cytotrophoblast cells belongs to the LDL-receptor protein  
RT superfamily.";  
RL EXP. CELL RES. 212:344-350(1994).  
RN [5]  
RP FUNCTION.  
RX MEDLINE; 95286588.  
RA KOUNNAS M.Z., LOUKINOVA E.B., STEFANSSON S., HARMONY J.A.K.,  
RA BREWER B.H., STRICKLAND D.K., ARGRAVES W.S.;  
RT "Identification of glycoprotein 330 as an endocytic receptor for  
RT apolipoprotein J/glycusterin.";  
RL J. BIOL. CHEM. 270:13070-13075(1995).  
CC -!- FUNCTION: BINDS SPECIFICALLY CLUSTERIN WITH HIGH AFFINITY, BUT  
CC ALSO LIGANDS IN COMMON WITH OTHER FAMILY MEMBERS: PLASMINOGEN,  
CC EXTRACELLULAR MATRIX COMPONENTS, PLASMINOGEN ACTIVATOR-PLASMINOGEN  
CC ACTIVATOR INHIBITOR TYPE I COMPLEX, APOLIPOPROTEIN E-ENRICHED  
CC BETA-VLDL, LIPOPROTEIN LIPASE, LACTOFERRIN AND CALCITUM.  
CC -!- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS  
CC APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B (BY SIMILARITY).  
CC -!- FUNCTION: MAY PARTICIPATE IN REGULATION OF PARATHYROID-HORMONE AND  
CC PARA-THYROID-HORMONE-RELATED PROTEIN RELEASE.  
CC -!- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-  
CC ASSOCIATED PROTEIN (RAP).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: ABSORPTIVE EPITHELIA, INCLUDING RENAL  
CC PROXIMAL TUBULES.  
CC -!- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.  
CC -!- SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS.  
CC -!- SIMILARITY: CONTAINS 17 EGF-LIKE REPEATS.  
CC -!- SIMILARITY: CONTAINS 4 SH2-BINDING DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 SH3-BINDING DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U33837; G1809240; -  
DR EMBL; U04441; G1389559; -  
DR EMBL; S73145; G685061; -  
DR MIM; 600073; -  
DR PROSITE; PS00010; ASX\_HYDROXYL; 4.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 9.  
DR PROSITE; PS01187; EGF\_CA; 3.  
DR PROSITE; PS01209; LDLRA\_1; 31.  
DR PROSITE; PS00068; LDLRA\_2; 36.  
DR PFAM; PF00057; ldl\_recept\_a; 36.  
DR PFAM; PF00058; ldl\_recept\_b; 37.  
DR HSP; P01130; IAUJ.  
KW GLYCOPROTEIN; REPEAT; ENDOCYTOSIS; COATED PITS; TRANSMEMBRANE;  
KW RECEPTOR; EGF-LIKE DOMAIN; SIGNAL; POLYMORPHISM.  
FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 4655  
FT DOMAIN 26 4423  
FT TRANSMEM 4424 4446  
FT DOMAIN 4447 4655  
FT DOMAIN 26 64  
FT DOMAIN 65 105  
FT DOMAIN 106 144  
FT DOMAIN 145 181  
FT DOMAIN 182 219  
FT DOMAIN 220 258  
FT DOMAIN 264 308  
FT DOMAIN 309 347  
FT DOMAIN 348 386  
FT DOMAIN 436 477  
FT DOMAIN 479 520  
FT DOMAIN 522 567  
FT DOMAIN 569 612  
FT DOMAIN 613 653  
FT DOMAIN 659 705  
FT DOMAIN 753 794  
FT DOMAIN 796 836  
FT DOMAIN 838 880  
FT DOMAIN 882 924  
FT DOMAIN 970 1014  
FT DOMAIN 1024 1062  
FT DOMAIN 1065 1103  
FT DOMAIN 1107 1145  
FT DOMAIN 1147 1185  
FT DOMAIN 1186 1224  
FT DOMAIN 1228 1268  
FT DOMAIN 1269 1307  
FT DOMAIN 1310 1350  
FT DOMAIN 1349 1389  
FT DOMAIN 1390 1429  
FT DOMAIN 1478 1519  
FT DOMAIN 1521 1562  
FT DOMAIN 1566 1608  
FT DOMAIN 1610 1653  
FT DOMAIN 1655 1695  
FT DOMAIN 1700 1741  
FT DOMAIN 1790 1831  
FT DOMAIN 1833 1881  
FT DOMAIN 1883 1929  
FT DOMAIN 1931 1971  
FT DOMAIN 1972 2012  
FT DOMAIN 2018 2059  
FT DOMAIN 2107 2155  
FT DOMAIN 2157 2200  
FT DOMAIN 2202 2244  
FT DOMAIN 2246 2288  
FT DOMAIN 2290 2331  
FT DOMAIN 2342 2383  
FT DOMAIN 2431 2476  
FT DOMAIN 2478 2517  
FT DOMAIN 2519 2561  
FT DOMAIN 2563 2603  
FT DOMAIN 2604 2646  
FT DOMAIN 2651 2693  
FT DOMAIN 2698 2738  
FT DOMAIN 2739 2777  
FT DOMAIN 2778 2819  
FT DOMAIN 2820 2861  
FT DOMAIN 2862 2901  
FT DOMAIN 2904 2945  
FT DOMAIN 2946 2990  
FT DOMAIN 2991 3029  
FT DOMAIN 3030 3070  
FT DOMAIN 3073 3110  
FT DOMAIN 3111 3151  
FT DOMAIN 3152 3192  
FT DOMAIN 3239 3280  
FT DOMAIN 3282 3331

LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED  
PROTEIN 2.  
EXTRACELLULAR (POTENTIAL).  
CYTOPLASMIC (POTENTIAL).  
LDL-RECEPTOR CLASS A 1.  
LDL-RECEPTOR CLASS A 2.  
LDL-RECEPTOR CLASS A 3.  
LDL-RECEPTOR CLASS A 4.  
LDL-RECEPTOR CLASS A 5.  
LDL-RECEPTOR CLASS A 6.  
LDL-RECEPTOR CLASS A 7.  
EGF-LIKE 1.  
EGF-LIKE 2.  
LDL-RECEPTOR CLASS B 1.  
LDL-RECEPTOR CLASS B 2.  
LDL-RECEPTOR CLASS B 3.  
LDL-RECEPTOR CLASS B 4.  
LDL-RECEPTOR CLASS B 5.  
EGF-LIKE 3.  
LDL-RECEPTOR CLASS B 6.  
LDL-RECEPTOR CLASS B 7.  
LDL-RECEPTOR CLASS B 8.  
LDL-RECEPTOR CLASS B 9.  
EGF-LIKE 4.  
LDL-RECEPTOR CLASS A 8.  
LDL-RECEPTOR CLASS A 9.  
LDL-RECEPTOR CLASS A 10.  
LDL-RECEPTOR CLASS A 11.  
LDL-RECEPTOR CLASS A 12.  
LDL-RECEPTOR CLASS A 13.  
LDL-RECEPTOR CLASS A 14.  
LDL-RECEPTOR CLASS A 15.  
EGF-LIKE 5.  
EGF-LIKE 6.  
CALCIUM-BINDING (POTENTIAL).  
LDL-RECEPTOR CLASS B 10.  
LDL-RECEPTOR CLASS B 11.  
LDL-RECEPTOR CLASS B 12.  
LDL-RECEPTOR CLASS B 13.  
LDL-RECEPTOR CLASS B 14.  
EGF-LIKE 7.  
LDL-RECEPTOR CLASS B 15.  
LDL-RECEPTOR CLASS B 16.  
LDL-RECEPTOR CLASS B 17.  
LDL-RECEPTOR CLASS B 18.  
LDL-RECEPTOR CLASS B 19.  
EGF-LIKE 8.  
LDL-RECEPTOR CLASS B 20.  
LDL-RECEPTOR CLASS B 21.  
LDL-RECEPTOR CLASS B 22.  
LDL-RECEPTOR CLASS B 23.  
LDL-RECEPTOR CLASS B 24.  
EGF-LIKE 9.  
LDL-RECEPTOR CLASS B 25.  
LDL-RECEPTOR CLASS B 26.  
LDL-RECEPTOR CLASS B 27.  
LDL-RECEPTOR CLASS B 28.  
LDL-RECEPTOR CLASS B 29.  
EGF-LIKE 10.  
LDL-RECEPTOR CLASS A 16.  
LDL-RECEPTOR CLASS A 17.  
LDL-RECEPTOR CLASS A 18.  
LDL-RECEPTOR CLASS A 19.  
LDL-RECEPTOR CLASS A 20.  
LDL-RECEPTOR CLASS A 21.  
LDL-RECEPTOR CLASS A 22.  
LDL-RECEPTOR CLASS A 23.  
LDL-RECEPTOR CLASS A 24.  
LDL-RECEPTOR CLASS A 25.  
EGF-LIKE 11.  
EGF-LIKE 12.  
CALCIUM-BINDING (POTENTIAL).  
LDL-RECEPTOR CLASS B 30.  
LDL-RECEPTOR CLASS B 31.

FT DOMAIN 3333 3375 LDL-RECEPTOR CLASS B 32.  
 FT DOMAIN 3377 3418 LDL-RECEPTOR CLASS B 33.  
 FT DOMAIN 3419 3459 LDL-RECEPTOR CLASS B 34.  
 FT DOMAIN 3465 3509 EGF-LIKE 13.  
 FT DOMAIN 3510 3550 LDL-RECEPTOR CLASS A 26.  
 FT DOMAIN 3551 3591 LDL-RECEPTOR CLASS A 27.  
 FT DOMAIN 3592 3632 LDL-RECEPTOR CLASS A 28.  
 FT DOMAIN 3633 3673 LDL-RECEPTOR CLASS A 29.  
 FT DOMAIN 3676 3716 LDL-RECEPTOR CLASS A 30.  
 FT DOMAIN 3717 3756 LDL-RECEPTOR CLASS A 31.  
 FT DOMAIN 3757 3795 LDL-RECEPTOR CLASS A 32.  
 FT DOMAIN 3796 3834 LDL-RECEPTOR CLASS A 33.  
 FT DOMAIN 3840 3880 LDL-RECEPTOR CLASS A 34.  
 FT DOMAIN 3881 3922 LDL-RECEPTOR CLASS A 35.  
 FT DOMAIN 3926 3964 EGF-LIKE 14.  
 FT DOMAIN 3966 4006 EGF-LIKE 15.  
 FT DOMAIN 4007 4048 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 4154 4195 LDL-RECEPTOR CLASS B 35.  
 FT DOMAIN 4197 4239 LDL-RECEPTOR CLASS B 36.  
 FT DOMAIN 4242 4282 LDL-RECEPTOR CLASS B 37.  
 FT DOMAIN 4330 4368 EGF-LIKE 16.  
 FT DOMAIN 4377 4411 EGF-LIKE 17.  
 FT SITE 1742 1744 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 4526 4526 CRITICAL FOR ENDOCYTOSIS (POTENTIAL).  
 FT SITE 4598 4598 CRITICAL FOR ENDOCYTOSIS (POTENTIAL).  
 FT DISULFID 28 40 BY SIMILARITY.  
 FT DISULFID 35 53 BY SIMILARITY.  
 FT DISULFID 47 62 BY SIMILARITY.  
 FT DISULFID 67 80 BY SIMILARITY.  
 FT DISULFID 74 93 BY SIMILARITY.  
 ...  
 Note: remainder of annotations omitted.

Query Match 67.6%; Score 46; DB 1: Length 4655;  
 Best Local Similarity 50.0%; Pred. No. 5.66e+00;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 261 HDVHCSP 268  
 I : I : I :  
 QY 481 HAIHRCHP 488

RESULT 12  
 ID VARY\_YEAST STANDARD; PRT; 345 AA.  
 AC P32366;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE VACUOLAR ATP SYNTHASE SUBUNIT AC39 (EC 3.6.1.34) (V-ATPASE AC39 SUBUNIT) (V-ATPASE SUBUNIT M39).  
 GN VMA6 OR YLR447C OR YLR324.8.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMITASCOMYCETES; SACCCHAROMYCETALES;  
 CC SACCCHAROMYCETACEAE; SACCCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE; 93286119.  
 RA BAUERLE C., HO M.N., LINDORFER M.A., STEVENS T.H.;  
 RT "The Saccharomyces cerevisiae VMA6 gene encodes the 36-kDa subunit of the vacuolar H(+)-ATPase membrane sector.";  
 RL J. BIOL. CHEM. 268:12749-12757(1993).  
 CC -!- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS. THE ACTIVE ENZYME CONSISTS OF A CATALYTIC V1 DOMAIN ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE COMPLEX. THIS SUBUNIT IS A NON-INTEGRAL MEMBRANE COMPONENT OF THE MEMBRANE PORE DOMAIN AND IS REQUIRED FOR PROPER ASSEMBLY OF THE V0 SECTOR. MIGHT BE INVOLVED IN THE REGULATED ASSEMBLY OF V1 SUBUNITS ONTO THE MEMBRANE SECTOR OR ALTERNATIVELY MAY PREVENT THE PASSAGE OF PROTONS THROUGH V0 PORES.  
 CC -!- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.  
 CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS V-ATPASE SUBUNIT AC39, V-TYPE SODIUM ATPASE NTPC AND ARCHEAL ATPASE SUBUNIT C.  
 -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL; L11584; G173171; -  
 DR EMBL; U22382; G717067; -  
 DR PIR; S35105; S35105;  
 DR PIR; A45994; A45994.  
 DR SGD; L0002461; VMA6.  
 KW HYDROLASE; HYDROGEN ION TRANSPORT.  
 FT CONFLICT 32 32 N -> T (IN REF. 1).  
 SQ SEQUENCE 345 AA; 39790 MW; 76C2B483 CRC32;  
 -----  
 Query Match 66.2%; Score 45; DB 1: Length 345;  
 Best Local Similarity 66.7%; Pred. No. 9.16e+00;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 124 LORCHP 129  
 I : I : I : I :  
 QY 483 IHRCHP 488

RESULT 13  
 ID BIND\_STREN STANDARD; PRT; 485 AA.  
 AC P23118;  
 DT 01-NOV-1991 (REL. 20, CREATED)  
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE BINDIN PRECURSOR.  
 OS STRONGYLOCENTROTUS FRANCISCANUS (SEA URCHIN).  
 OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;  
 CC EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92130802.  
 RA MINOR J.E., FROMSON D.R., BRITTEN R.J., DAVIDSON E.H.;  
 RT "Comparison of the bindin proteins of Strongylocentrotus franciscanus, S. purpuratus, and Lytechinus variegatus: sequences involved in the species specificity of fertilization.";  
 RL MOL. BIOL. EVOL. 8:781-795(1991).  
 RN [2]  
 RP PRELIMINARY SEQUENCE OF 248-320.  
 RA VACQUIER V.D., MOY G.W.;  
 RL (IN) DIRKSEN E.R., PRESCOTT D., FOX C.F. (EDS.);  
 CC CELL REPRODUCTION, PP.12:379-389, ACADEMIC PRESS, NEW YORK (1978).  
 CC -!- FUNCTION: SPECIES-SPECIFIC SEA URCHIN SPERM PROTEIN REQUIRED FOR ADHESION OF SPERM TO THE EGG SURFACE DURING FERTILIZATION. BINDIN COATS THE ACROSOMAL PROCESS AFTER IT IS EXTERNALIZED BY THE ACROSOME REACTION. IT BINDS TO SULFATED, FUCOSE-CONTAINING POLYSACCHARIDES ON THE VITELLINE LAYER RECEPTOR PROTEOGLYCAN WHICH COVER THE EGG PLASMA MEMBRANE.  
 CC -!- SUBCELLULAR LOCATION: MAJOR PROTEIN COMPONENT OF THE LUMEN OF THE ACROSOME GRANULE.  
 CC -!- SIMILARITY: TO OTHER SEA URCHINS BINDIN.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL; M59490; G161443; -  
 DR PIR; B40552; B40552.  
 KW SPERM; SIGNAL; FERTILIZATION.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT PROPEP 21 247 POTENTIAL.  
 -----

FT CHAIN 248 485 BINDIN.  
FT DOMAIN 371 379 FUCOSE-BINDING DOMAIN (POTENTIAL).  
FT DOMAIN 399 406 POLY-GLU.  
SQ SEQUENCE 485 AA; 51940 MW; 60193AAB CRC32;

Query Match 66.2%; Score 45; DB 1; Length 485;  
Best Local Similarity 80.0%; Pred. No. 9.16e+00;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 146 VHRHC 150  
:|||||  
QY 483 IHRCH 487

RESULT 14  
ID DPOL\_ADE05 STANDARD; PRT; 1056 AA.  
AC P04495;  
DT 13-AUG-1987 (REL. 05, CREATED)  
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
DT 13-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7).  
GN POL.  
OS HUMAN ADENOVIRUS TYPE 5.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84183604.  
RA DEKKER B.M.M., VAN ORMONDT H.;  
RT "The nucleotide sequence of fragment HindIII-C of human adenovirus  
type 5 DNA (map positions 17.1-31.7).";  
RL GENE 27:115-120(1984).  
CC -|- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -  
N PYROPHOSPHATE + DNA(N).  
CC -|- THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.  
CC -|- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X02996; G58495; -  
CC PIR; A00712; DJAD51.  
CC PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
CC TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION;  
CC DNA-BINDING.  
CC SEQUENCE 1056 AA; 120400 MW; 822A9B85 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 1056;  
Best Local Similarity 42.9%; Pred. No. 9.16e+00;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 81 QPVHRCQ 87  
:|||||  
QY 481 HAIHRCH 487

RESULT 15  
ID DPOL\_ADE02 STANDARD; PRT; 1056 AA.  
AC P03261;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 13-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE DNA POLYMERASE (EC 2.7.7.7).  
GN POL.  
OS HUMAN ADENOVIRUS TYPE 2.  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 83056843.

RA GINGERAS T.R., SCIACKY D., GELINAS R.E., BING-DONG J., YEN C.E.,  
RA KELLY M.M., BULLOCK P.A., PARSONS B.L., O'NEILL K.E., ROBERTS R.J.;  
RT "Nucleotide sequences from the adenovirus-2 genome.";  
RL J. BIOL. CHEM. 257:13475-13491(1982).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 83056844.  
RA ALSTROM P., AKUSJARVI G., PETERSSON M., PETERSSON U.;  
RT "DNA sequence analysis of the region encoding the terminal protein  
and the hypothetical N-gene product of adenovirus type 2.";  
RL J. BIOL. CHEM. 257:13492-13498(1982).  
CC -|- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -  
N PYROPHOSPHATE + DNA(N).  
CC -|- THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.  
CC -|- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; J01917; G209822; -  
CC PIR; A00711; WMAD12.  
CC PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
CC TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION;  
CC DNA-BINDING.  
CC SEQUENCE 1056 AA; 120432 MW; C53DE5E4 CRC32;  
Query Match 66.2%; Score 45; DB 1; Length 1056;  
Best Local Similarity 42.9%; Pred. No. 9.16e+00;  
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 81 QPVHRCQ 87  
:|||||  
QY 481 HAIHRCH 487

Search completed: Thu May 20 13:19:13 1999  
Job time : 12 secs.

\*\*\*\*\*

W P E R E L A (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 20 13:19:31 1999; MasPar time 5.62 Seconds

Tabular output not generated. 77.652 Million cell updates/sec

Title: >US-09-099-053-2  
Description: (481-488) from US0909053.pep (6 of 6)  
Perfect Score: 68  
Sequence: 1 HAIHRCHP 8

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb19  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 20.687; Variance 24.857; scale 0.832

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	72.1	143	2	Q50818	ORF IS HOMOLOGOUS TO U
2	49	72.1	346	5	O02240	MEC-3 PROTEIN.
3	49	72.1	360	5	O01657	FRUCTOSE-BISPHOSPHATE
4	49	72.1	597	5	O45218	ALKYL-DIHYDROXYACETONE
5	49	72.1	3414	14	O88493	POLYPROTEIN.
6	49	72.1	3414	14	O10383	POLYPROTEIN.
7	49	72.1	3414	14	O88489	POLYPROTEIN.
8	47	69.1	393	5	O09487	HYPOTHETICAL 45.4 KD P
9	47	69.1	470	13	O93389	SPADETAIL.
10	47	69.1	470	13	O93303	T-BOX CONTAINING PROTE
11	47	69.1	705	10	O04620	SIMILARITY TO GALECTIN
12	47	69.1	897	14	O96631	PROTEIN A.
13	47	69.1	932	5	O01623	SIMILAR TO LIGAND-GATE
14	47	69.1	998	14	O66929	PROTEIN A.
15	46	67.6	513	5	O45243	F57C2.6 PROTEIN.
16	46	67.6	581	14	O84694	65KDA VIRAL REPLICASE
17	46	67.6	1046	5	O44316	REVERSE TRANSCRIPTASE
18	45	66.2	95	14	O69132	EPSTEIN-BARR VIRUS LAT
19	45	66.2	147	10	O23990	EXPRESSED SEQUENCE TAG
20	45	66.2	178	11	O08584	CORE PROMOTER BINDING

21	45	66.2	215	4	O43839	BCD ORF2.	1.42e+01
22	45	66.2	216	9	O80201	HYPOTHETICAL 25.3 KD P	1.42e+01
23	45	66.2	253	5	O45560	F54B8.2 PROTEIN.	1.42e+01
24	45	66.2	255	5	P91575	W05B10.4 PROTEIN.	1.42e+01
25	45	66.2	283	11	O35819	ZINC FINGER PROTEIN.	1.42e+01
26	45	66.2	290	4	O39612	DNA-BINDING PROTEIN CP	1.42e+01
27	45	66.2	1194	3	O42649	HYPOTHETICAL 136.8 KD	1.42e+01
28	45	66.2	2044	5	P91620	STILL LIFE TYPE 2.	1.42e+01
29	44	64.7	751	4	O99985	SEMAPHORIN E.	2.30e+01
30	44	64.7	751	11	Q62181	SEMAPHORIN E PRECURSOR	2.30e+01
31	44	64.7	761	13	Q90663	COLLAPSIDIN-2.	2.30e+01
32	44	64.7	796	2	O25163	BIOTIN SULFOXIDE REDUC	2.30e+01
33	44	64.7	1030	4	O76039	STK9 PROTEIN.	2.30e+01
34	44	64.7	1106	5	O18633	C44H9.4 PROTEIN.	2.30e+01
35	44	64.7	1125	5	O93203	CL1E4.6 PROTEIN.	2.30e+01
36	44	64.7	1461	2	O05819	HYPOTHETICAL 156.7 KD	2.30e+01
37	44	64.7	2796	2	Q48926	FATTY ACID SYNTHASE.	2.30e+01
38	44	64.7	4545	11	Q61291	LOW DENSITY LIPOPROTEI	2.30e+01
39	43	63.2	292	14	O36377	ORF27.	3.71e+01
40	43	63.2	364	14	O83833	MATRIX PROTEIN.	3.71e+01
41	43	63.2	389	5	Q24276	CELL DIVISION CYCLE 37	3.71e+01
42	43	63.2	730	10	O22989	CELLULOSE SYNTHASE ISO	3.71e+01
43	43	63.2	815	11	O88839	METALLOPROTEASE-DISINT	3.71e+01
44	43	63.2	1313	3	P87141	PUTATIVE GUANINE-NUCLE	3.71e+01
45	43	63.2	1548	11	Q62040	PC6B (FRAGMENT).	3.71e+01

ALIGNMENTS

RESULT 1  
ID Q50818 PRELIMINARY; PRT; 143 AA.  
AC Q50818;  
DT 01-NOV-1996 (TREMUREL. 01, CREATED)  
DT 01-NOV-1996 (TREMUREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMUREL. 07, LAST ANNOTATION UPDATE)  
DE ORF IS HOMOLOGOUS TO UNIDENTIFIED ORF ADJACENT TO DNAB IN BACILLUS  
DE SUBTILIS.  
OS MYCOBACTERIUM TUBERCULOSIS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=50410;  
RA PATKI A.H.; DALE J.W.;  
RL SUBMITTED (APR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; X59273; E36552; .  
SQ SEQUENCE 143 AA; 15414 MW; FC3BA2F0 CRC32;

Query Match 72.1%; Score 49; DB 2; Length 143;  
Best Local Similarity 71.4%; Pred. No. 1.92e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 104 ALQRCHP 110  
Qy 482 AIHRCHP 488

RESULT 2  
ID O02240 PRELIMINARY; PRT; 346 AA.  
AC O02240;  
DT 01-JUL-1997 (TREMUREL. 04, CREATED)  
DT 01-JAN-1998 (TREMUREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMUREL. 07, LAST ANNOTATION UPDATE)  
DE MEC-3 PROTEIN.  
GN MEC-3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WILD A.;  
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KRISTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL NATURE 368:32-38(1994).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL: Z81054; E1345139; -.  
DR PROSITE: PS00027; HOMEBOX\_1; 1.  
KW HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN.  
SQ SEQUENCE 346 AA; 39989 MW; 71016FFF CRC32;  
  
Query Match 72.1%; Score 49; DB 5; Length 346;  
Best Local Similarity 83.3%; Pred. No. 1.92e+00;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Db 109 HSIHRC 114  
Qy 481 HAIHRC 486  
I:||||  
  
RESULT 3  
ID O01657; PRELIMINARY; PRT; 360 AA.  
AC O01657;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE FRUCTOSE-BISPHOSPHATE ALDOLASE (FRAGMENT).  
OS ONCHOCERCA VOLVULUS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; SPIRURIA; SPIRURIDA;  
OC FILARIOIDEA; ONCHOCERCIDAE; ONCHOCERCA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FOREST;  
RA JOSEPH G.T., HUIMA T., LUSTIGMAN S.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U96178; G1945479; -.  
DR PROSITE: PS00158; ALDOLASE\_CLASS\_I; 1.  
DR PFAM: PF00274; glycolytic\_enzy; 1.  
FT NON\_TER 1  
SQ SEQUENCE 360 AA; 39166 MW; 090B2E33 CRC32;  
  
Query Match 72.1%; Score 49; DB 5; Length 360;  
Best Local Similarity 57.1%; Pred. No. 1.92e+00;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 194 HDVHRC 200  
Qy 481 HAIHRC 487  
I:||||  
  
RESULT 4  
ID O45218; PRELIMINARY; PRT; 597 AA.  
AC O45218;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE ALKYL-DIHYDROXYACETONEPHOSPHATE SYNTHASE (EC 2.5.1.26)  
DE (ALKYLGLYCERONE-PHOSPHATE SYNTHASE)  
DE (ALKYLDIHYDROXYACETONEPHOSPHATE SYNTHASE).  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-CB1489 HIM-8(E1489); TISSUE-WHOLE ANIMAL;  
RX MEDLINE: 98113342.  
RA DE VET E.C.J.M., PRINSEN H.C.M.T., VAN DEN BOSCH H.;  
RT "Nucleotide sequence of a cDNA clone encoding a Caenorhabditis  
RT elegans homolog of mammalian alkyl-dihydroxyacetonephosphate  
RT synthase: evolutionary switching of peroxisomal targeting signals.";  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 242:277-281(1998).  
CC -!- CATALYTIC ACTIVITY: 1-ACYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN  
CC ALCOHOL = 1-ALKYL-GLYCERONE 3-PHOSPHATE + A LONG-CHAIN ACID  
CC ANION.  
DR EMBL: AJ002686; E1248248; -.  
KW TRANSFERASE.  
SQ SEQUENCE 597 AA; 66559 MW; AFFB98EE CRC32;  
  
Query Match 72.1%; Score 49; DB 5; Length 597;  
Best Local Similarity 71.4%; Pred. No. 1.92e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 338 AIORCQP 344  
Qy 482 AIHRC 488  
I:||||  
  
RESULT 5  
ID Q88493; PRELIMINARY; PRT; 3414 AA.  
AC Q88493;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE POLYPROTEIN.  
OS TICK-BORNE ENCEPHALITIS VIRUS (TBEV).  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;  
OC FLAVIVIRUS.  
RN [1]  
RP SEQUENCE OF 1-790 FROM N.A.  
RC STRAIN-NEUDOERFL;  
RX MEDLINE: 88322870.  
RA MANDL C.W., HEINZ F.X., KUNZ C.;  
RT "Sequence of the structural proteins of tick-borne encephalitis virus  
RT (western subtype) and comparative analysis with other flaviviruses.";  
RL VIROLOGY 166:197-205(1988).  
RN [2]  
RP SEQUENCE OF 773-3414 FROM N.A.  
RC STRAIN-NEUDOERFL;  
RX MEDLINE: 90051080.  
RA MANDL C.W., HEINZ F.X., STOCKL E., KUNZ C.;  
RT "Genome sequence of tick-borne encephalitis virus (Western subtype)  
RT and comparative analysis of nonstructural proteins with other  
RL flaviviruses.";  
RL VIROLOGY 173:291-301(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NEUDOERFL;  
RX MEDLINE: 96036491.  
RA WALLNER G., MANDL C.W., KUNZ C., HEINZ F.X.;  
RT "The flavivirus 3'-noncoding region: extensive size heterogeneity  
RT independent of evolutionary relationships among strains of tick-borne  
RL encephalitis virus.";  
RL VIROLOGY 213:169-178(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NEUDOERFL;  
RA MANDL C.W.;  
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U27495; G975238; -.  
DR PFAM: PF00869; Flavi\_glycoprot; 1.  
DR PFAM: PF00948; Flavi\_NSI; 1.  
DR PFAM: PF00949; Flavi\_helicase; 1.  
DR PFAM: PF00972; Flavi\_NS5; 1.  
DR PFAM: PF01002; Flavi\_NS2B; 1.  
DR PFAM: PF01003; Flavi\_capsid; 1.  
DR PFAM: PF01004; Flavi\_M; 1.

```
DR PFAM; PF01005; Flavi_NS2A; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3414 AA; 378320 MW; 9CA32E8A CRC32;

Query Match
Best Local Similarity 72.1%; Score 49; DB 14; Length 3414;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HEVHQCOP 1604
| :|:|:|
QY 481 HAIHRCHP 488

RESULT 6
ID O10383 PRELIMINARY; PRT; 3414 AA.
AC O10383;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS LOUPING ILL VIRUS (LI).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=369/T2;
RX MEDLINE; 97321792.
RA GRITSUN T.S.; VENUGOPAL K.; DE ZANOTTO P.M.; MIKHAILOV M.V.;
RA SALL A.A.; POLKINGHORNE I.; FROLOVA T.V.; POGODINA V.V.;
RA LASHKEVICH V.A.; GOULD E.A.;
RT "Complete sequence of two tick-borne flaviviruses isolated from
RT Siberia and the UK: analysis and significance of the 5' and
RT 3'-UTRs.";
RL VIRUS RES. 49:27-39(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=369/T2;
RA GRITSUN T.S.;
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Y07863; E281315; -.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
KW POLYPROTEIN; ENVELOPE PROTEIN; MEMBRANE.
FT CHAIN 1 112 CAPSID (CORE) PROTEIN.
FT CHAIN 113 280 PREMEMBRANE PROTEIN.
FT CHAIN 206 280 MEMBRANE PROTEIN.
FT CHAIN 281 776 ENVELOPE PROTEIN.
FT CHAIN 777 1128 ENVELOPE PROTEIN.
FT CHAIN 1129 1358 NONSTRUCTURAL PROTEIN.
FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN.
FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN.
FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN.
FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN.
FT CHAIN 2512 3414 RNA POLYMERASE.
SQ SEQUENCE 3414 AA; 378410 MW; 7DOA5DDE CRC32;

Query Match
Best Local Similarity 72.1%; Score 49; DB 14; Length 3414;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HEVHQCOP 1604
| :|:|:|
QY 481 HAIHRCHP 488

RESULT 7
ID Q88489 PRELIMINARY; PRT; 3414 AA.
AC Q88489;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS TICK-BORNE ENCEPHALITIS VIRUS (TBEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=263;
RX MEDLINE; 96036491.
RA WALLNER G.; MANDL C.W.; KUNZ C.; HEINZ F.X.;
RT "The flavivirus 3'-noncoding region: extensive size heterogeneity
RT independent of evolutionary relationships among strains of tick-borne
RT encephalitis virus.";
RL VIROLOGY 213:169-178(1995).
DR EMBL; U27491; G1063399; -.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3414 AA; 378064 MW; 03C4BD6C CRC32;

Query Match
Best Local Similarity 72.1%; Score 49; DB 14; Length 3414;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1597 HEVHQCOP 1604
| :|:|:|
QY 481 HAIHRCHP 488

RESULT 8
ID Q09487 PRELIMINARY; PRT; 393 AA.
AC Q09487;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 45.4 KD PROTEIN C30G12.7 IN CHROMOSOME II.
GN C30G12.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITIA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA LATREILLE P.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U21319; G687839; -.
DR WORMPEP; C30G12.7; CE01838.
DR PFAM; PF00806; PUF; 8.
KW HYPOTHETICAL PROTEIN; REPEAT.
FT DOMAIN 51 271 6 APPROXIMATE TANDEM REPEATS.
FT REPEAT 51 88 1.
FT REPEAT 89 124 2.
FT REPEAT 125 160 3.
FT REPEAT 161 196 4.
FT REPEAT 197 235 5.
FT REPEAT 236 271 6.
SQ SEQUENCE 393 AA; 45433 MW; AD83C19E CRC32;

Query Match
Best Local Similarity 69.1%; Score 47; DB 5; Length 393;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 164 HVIHRC 169
| | | | |
```

QY 481 HAIHRC 486

RESULT 9

ID O93389 PRELIMINARY; PRT; 470 AA.

AC O93389;

DT 01-NOV-1998 (TREMREL. 08, CREATED)

DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)

DE SPADETAIL.

GN SPT.

OS BRACHYDANTIO RERIO (ZEBRAFISH) (ZEBRA DANIO).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;

OC TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;

OC CYPRINIDAE; RASBORINAE; DANIO.

RN [1]

RP SEQUENCE FROM N.A.

RA GRIFFIN K.J.P., AMACHER S.L., KIMMEL C.B., KIMELMAN D.;

RT "Molecular identification of spadetail: regulation of zebrafish trunk and tail formation by T-box genes.";

RL DEVELOPMENT 125:0-0(1998).

DR EMBL: AF077225; G3396048;

DR PROSITE: PS01264; TBOX.2; 1.

DR PROSITE: PS01283; TBOX.1; 1.

SQ SEQUENCE 470 AA; 52579 MW; F76B4182 CRC32;

Query Match 69.1%; Score 47; DB 13; Length 470;

Best Local Similarity 62.5%; Pred. No. 5.29e+00;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 151 HSMHYHP 158

QY 481 HAIHRC 488

RESULT 10

ID O93303 PRELIMINARY; PRT; 470 AA.

AC O93303;

DT 01-NOV-1998 (TREMREL. 08, CREATED)

DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)

DE T-BOX CONTAINING PROTEIN.

GN TBX16.

OS BRACHYDANTIO RERIO (ZEBRAFISH) (ZEBRA DANIO).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;

OC TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;

OC CYPRINIDAE; RASBORINAE; DANIO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 98237855.

RA RUVINSKY I., SILVER L.M., HO R.K.;

RT "Characterization of the zebrafish tbx16 gene and evolution of the vertebrate T-box family.";

RL DEV. GENES EVOL. 208:94-99(1998).

DR EMBL: AF044977; G3273697;

DR PROSITE: PS01264; TBOX.2; 1.

DR PROSITE: PS01283; TBOX.1; 1.

SQ SEQUENCE 470 AA; 52618 MW; 74BBB7AE CRC32;

Query Match 69.1%; Score 47; DB 13; Length 470;

Best Local Similarity 62.5%; Pred. No. 5.29e+00;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 151 HSMHYHP 158

QY 481 HAIHRC 488

RESULT 11

ID O04620 PRELIMINARY; PRT; 705 AA.

AC O04620;

DT 01-JUL-1997 (TREMREL. 04, CREATED)

DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)

DE SIMILARITY TO GALECTIN-3 BINDING PROTEIN.

GN A\_IG02N01.13.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).

OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;

OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;

OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA SCHEET P., MAGGI L.;

RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA WATERSTON R.;

RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AF007269; G2191134;

SQ SEQUENCE 705 AA; 78735 MW; 6559C0CE CRC32;

Query Match 69.1%; Score 47; DB 10; Length 705;

Best Local Similarity 57.1%; Pred. No. 5.29e+00;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 306 GYRRCHP 312

QY 482 AIHRC 488

RESULT 12

ID Q96631 PRELIMINARY; PRT; 897 AA.

AC Q96631;

DT 01-FEB-1997 (TREMREL. 02, CREATED)

DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)

DE PROTEIN A.

OS BLACK BEETLE VIRUS (BBV).

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NODAVIRIDAE;

OC NODAVIRUS.

RN [1]

RP SEQUENCE OF 893-897 FROM N.A.

RA GUARINO L.A., GHOSH A., DASMAHAPATRA B., DASGUPTA R., KAESBERG P.;

RL VIROLOGY 139:190-203(1984).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 85210903.

RA DASMAHAPATRA B., DASGUPTA R., GHOSH A., KAESBERG P.;

RT "Structure of the black beetle virus genome and its functional implications.";

RL J. MOL. BIOL. 182:183-189(1985).

DR EMBL: K02560; G210670;

SQ SEQUENCE 897 AA; 101870 MW; 54B851CC CRC32;

Query Match 69.1%; Score 47; DB 14; Length 897;

Best Local Similarity 62.5%; Pred. No. 5.29e+00;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 257 HKIHC 264

QY 481 HAIHRC 488

RESULT 13

ID O01623 PRELIMINARY; PRT; 932 AA.

AC O01623;

DT 01-JUL-1997 (TREMREL. 04, CREATED)

DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)

DE SIMILAR TO LIGAND-GATED IONIC CHANNEL PROTEINS.

GN ZC196.7.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.





\*\*\*\*\*  
M P E R L A  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu May 20 16:27:55 1999; Maspar time 3568.62 Seconds  
Tabular output not generated. 1202.402 Million cell updates/sec

Title: >US-09-099-053-1  
Description: (1-1548) from US0909053.seq  
Perfect Score: 1548  
N.A. Sequence: 1 GCTCGCGGGCTCCCATGGCC.....CCACGCTCTGGGCTCCAGC 1548  
Comp: CGAGCGCCCGAGGTACCGG.....GGTTGCGAGACCCGAGGTGCG

Scoring table: TABLE\_jmetric  
Gap 60

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb158  
1:em\_ba1 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
7:em\_in 8:em\_ov 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
13:em\_pl 14:em\_ro 15:em\_sts 16:em\_v1

Database: genbank111  
17:gb\_ba1 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
22:gb\_in2 23:gb\_ov 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
28:gb\_pl2 29:gb\_pl1 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_v1

Statistics: Mean 9.259; Variance 3.020; scale 3.066

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	32	2.1	2393 32 MUSTSPK Mouse mRNA for tyrosin 3.91e-11
2	32	2.1	2560 32 MUSTSRM Mouse mRNA for sm (ty 3.91e-11
3	22	1.4	4062 32 HAMNEU Syrian golden hamster 3.40e-02
4	21	1.4	5423 21 DME132670 Drosophila melanogaste 2.11e-01
5	21	1.4	162815 20 AC008494 Drosophila melanogaste 2.11e-01
6	22	1.4	183180 19 AC004929 Homo sapiens clone DJ0 3.40e-02
7	20	1.3	211 32 MUSPTKV Mus musculus protein t 1.23e+00
8	20	1.3	1653 37 RERSVSRRC Rous sarcoma virus pra 1.23e+00
9	20	1.3	1926 29 HUMHCKB Human hemopoietic cell 1.23e+00
10	20	1.3	2015 29 HUMHCKB Human hemopoietic cell 1.23e+00
11	20	1.3	2167 30 HSHCKE11 H.sapiens HCK gene for 1.23e+00
12	20	1.3	2526 24 XHCFFYN Xiphophorus c-fyn (Xf 1.23e+00
13	20	1.3	2545 37 RSVPRSRC Duck adapted Rous sarc 1.23e+00

14	20	1.3	2640 24 DRAJ5030 Danio rerio mRNA for E 1.23e+00
15	20	1.3	3123 32 RERSVH19 Hamster H-19 proviral 1.23e+00
16	20	1.3	3775 32 MAVSCR H.sapiens H-19 provira 1.23e+00
17	20	1.3	6354 29 HSDRD2INT M.sapiens DRD2 gene, i 1.23e+00
18	20	1.3	8516 24 TVIGSETS T.vulgaris DNA for rib 1.23e+00
19	20	1.3	9312 37 RSVSEQ Rous Sarcoma Virus (du 1.23e+00
20	20	1.3	9312 37 RERSV6 Rous sarcoma virus gen 1.23e+00
21	20	1.3	9392 37 AF033808 Rous sarcoma virus, co 1.23e+00
22	20	1.3	9625 37 ALRCG Rous sarcoma virus (Pr 1.23e+00
23	20	1.3	33654 31 AF050737 Homo sapiens dopamine 1.23e+00
24	20	1.3	17241 30 AC004014 Human BAC clone RG022K 1.23e+00
25	19	1.2	804 31 HSPETINS2 Homo sapiens RET proto 6.74e+00
26	19	1.2	1074 24 CHMYO Chicken mRNA for myoge 6.74e+00
27	19	1.2	1161 30 HSTHPKA H.sapiens mRNA cd33 fo 6.74e+00
28	19	1.2	1294 21 DMKRAKEN Drosophila melanogaste 6.74e+00
29	19	1.2	1416 30 AF028824 Homo sapiens Tax inter 6.74e+00
30	19	1.2	1685 30 HSU22171 Human DNA repair and r 6.74e+00
31	19	1.2	1765 31 AF089816 Homo sapiens RGS-GAIP 6.74e+00
32	19	1.2	1779 32 MMU05247 Mus musculus c-Src kin 6.74e+00
33	19	1.2	1849 17 SGSTRB1F S.galbus DSM 40480 gen 6.74e+00
34	19	1.2	1882 18 ACU40711 Aeromonas caviae hemol 6.74e+00
35	19	1.2	2015 17 SLFUSH S.lividans fush gene. 6.74e+00
36	19	1.2	2125 32 RPTVKI Rat mRNA for protein-t 6.74e+00
37	19	1.2	10446 17 AE000810 Methanobacterium therm 6.74e+00
38	19	1.2	31757 30 HSN53A9 Human DNA sequence fro 6.74e+00
39	19	1.2	43632 31 AC004144 Homo sapiens chromosom 6.74e+00
40	19	1.2	102389 21 AC004546 Drosophila melanogaste 6.74e+00
41	19	1.2	114324 20 AC000377 Human Chromosome 11 BA 6.74e+00
42	19	1.2	151187 19 HS460378 Human DNA sequence *** 6.74e+00
43	19	1.2	151965 20 AC006454 Homo sapiens clone DJ0 6.74e+00
44	19	1.2	157321 31 AC004912 Homo sapiens PAC clone 6.74e+00
45	19	1.2	192853 19 AC005096 Homo sapiens clone RG3 6.74e+00

ALIGNMENTS

RESULT 1  
LOCUS MUSTSPK 2393 bp mRNA ROD 06-FEB-1999  
DEFINITION Mouse mRNA for tyrosine-specific protein kinase, complete cds.  
ACCESSION D26186  
NID 9529072  
VERSION D26186.1 GI:529072  
KEYWORDS tyrosine-specific protein kinase.  
SOURCE Mus musculus adult lung cDNA to mRNA, clone\_lib:5' stretch.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2393)  
AUTHORS Kohmura, N.  
TITLE Direct Submission  
JOURNAL Submitted (24-DEC-1993) to the DDBJ/EMBL/GenBank databases. Naohiro Kohmura, National Institute for Physiological Sciences, Lab. Neurobiology & Behavioral Genetics; 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444, Japan (Tel:0564-55-7744, Fax:0564-55-7741)  
REFERENCE 2 (bases 1 to 2393)  
AUTHORS Kohmura, N., Yagi, T., Tomooka, Y., Oyanagi, M., Kominami, R., Takeda, N., Chiba, J., Ikawa, Y. and Aizawa, S.  
TITLE A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption  
JOURNAL Mol. Cell. Biol. 14 (10), 6915-6925 (1994)  
MEDLINE 95021220  
COMMENT Submitted (24-Dec-1993) to DDBJ by: Naohiro Kohmura  
Department of Neurobiology and Behavioral Genetics National Institute for Physiological Sciences  
Myodaiji, Okazaki 444  
Japan  
Phone: 0564-55-7741  
Fax: 0564-55-7744.  
Location/Qualifiers  
1. .2393  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

CDS

```

/genes="neu"
/product="putative"
33..3797
/genes="neu"
/codon_start=1
/product="p-185 precursor"
/protein_id="BAA03801.1"
/db_xref="PID:d1004317"
/db_xref="PID:g747595"
/db_xref="GI:747595"
/translation="MELAAWCGWGLLALLSPGASGQVCTGTDMLRLPASPETHLD
IVRLHVGCGVGVGQVLETLVLPANATLSFLQDIOEVGYMLIAHSQVRHVLPQLRLRV
RQTLQEDKVALVLDNRDPLDNTATGTPGLRELQLRLSLEILKGVLRGNFQ
LCQDILVMDKFRKNQLAPVDIDINRSCAPCPACDNHCWGPASDECQTLTGT
IAPRVAAPARARLPTDCHEQAGCTGPKHSDCLACLFHNSGICELHCPALVYNT
DTFSPNPGRTYFGASVYTCPTYNLSTEVGSLVCLPNNQEVTAEDGTQRCERK
SKSCARVYGLGHRARITSAIQEFAGCKKIFGSLAFIPESFDGPNSSGIAPL
TPQLOVFEETLEITGYLYTSAMPDSLDHLSVFQNLRAVIRGRVLHDGAYSILAQGLGI
RWGLRSLRLSGVLVLRHNTLHLCFVHTYPMQLFNPHQALLHNSNPEEGCLKD
FACIPCAHGHGPGPTQCVNCSHFURGQECVKECRWKGLPREYVNGKRLDPCPHE
CPONSTETGSEADQCTACPHYKDSFPCVAPCPGKPDLSYMPWKYPDEEGMOQ
PCPINTHSCVDLDERGCPAEORASPATSIATVVGILLFLVGVVGVGILIKRRQKI
RKYTMLRLQLEVEPTSGAMPNOAOMRLKETELRKVKVYLGSGAFGTVYKGIWI
PDGENVKIPVAIKVLENTSPKANKELIDEAVMAGLSGSPVSVLLCICLSTVQLVT
QMPYGCGLDVRHGRGLSGDQLLNMCVQIAKMSYLEDVRLVHOLAARNVLKSP
NHVKTIDFGARLLDDETEYHADGGKVPKIAIALESILRRRTHQSDVMSYGVYWE
LMTFGAKYDIPAREIPDLLEGERLPQPICTIDYIMVMKCMIDSECRPRFREL
VSFSRMARDPQFVYQIONEDLGPSSPLDSTFYRSLLEDMDGDLVDAEVLVPOQGF
FFPDAPAGAGTAHRRHSSSTRSGGELTLGMEPSGEEPRSPRLAPSEGAGSDVEEG
ELMGATKQPSIDRDLPLQRYSEDTPLPTEITDGYVAPLACSPQPEYVNGEYR
PQPLPEGLPFPVRPAGATLERKLTSPGKNGVYKDVFTFGGAVENTPEYLVPRGSSA
SQHPFALCPAFNLXYDDQDPSESGSPNTFTGTPAENPEYLGLDVVPY"
87..3794
/mat_peptide
/genes="neu"
/product="p-185"
516..1988
/misc_feature
/genes="neu"
/notes="extracellular domain"
1989..2057
/misc_feature
/genes="neu"
/notes="putative; transmembrane domain"
2058..3794
/misc_feature
/genes="neu"
/notes="cytoplasmic domain"
845 a 1209 c 1182 g 826 t
BASE COUNT 845 a 1209 c 1182 g 826 t
ORIGIN
Query Match 1.4%; Score 22; DB 32; Length 4062;
Best Local Similarity 100.0%; Pred. No. 3.40e-02;
Matches 22; Conservative 0; Mismatches --0; Indels --0; Gaps 0;
Db 2524 AGGCATGAGCTACCTGGAGGA 2545
|||||
QY 1055 AGGCATGAGCTACCTGGAGGA 1076

RESULT 4
LOCUS DME132670 5423 bp DNA INV 09-FEB-1999
DEFINITION Drosophila melanogaster buttonless gene, exon 1.
ACCESSION A0132670
NID 94544137
VERSION A0132670.1 GI:4454137
KEYWORDS buttonless gene; homeodomain; 1(3)L4910; P-element.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
1 (bases 1 to 5423)
Zhang, Y.Q.
Direct Submission
Submitted (06-FEB-1999) Zhang Y.Q., Biology, University of Utah,
257 South 1400 East, Salt Lake City, Utah 84112-0840, USA

/genes="neu"
/product="putative"
33..3797
/genes="neu"
/codon_start=1
/product="p-185 precursor"
/protein_id="BAA03801.1"
/db_xref="PID:d1004317"
/db_xref="PID:g747595"
/db_xref="GI:747595"
/translation="MELAAWCGWGLLALLSPGASGQVCTGTDMLRLPASPETHLD
IVRLHVGCGVGVGQVLETLVLPANATLSFLQDIOEVGYMLIAHSQVRHVLPQLRLRV
RQTLQEDKVALVLDNRDPLDNTATGTPGLRELQLRLSLEILKGVLRGNFQ
LCQDILVMDKFRKNQLAPVDIDINRSCAPCPACDNHCWGPASDECQTLTGT
IAPRVAAPARARLPTDCHEQAGCTGPKHSDCLACLFHNSGICELHCPALVYNT
DTFSPNPGRTYFGASVYTCPTYNLSTEVGSLVCLPNNQEVTAEDGTQRCERK
SKSCARVYGLGHRARITSAIQEFAGCKKIFGSLAFIPESFDGPNSSGIAPL
TPQLOVFEETLEITGYLYTSAMPDSLDHLSVFQNLRAVIRGRVLHDGAYSILAQGLGI
RWGLRSLRLSGVLVLRHNTLHLCFVHTYPMQLFNPHQALLHNSNPEEGCLKD
FACIPCAHGHGPGPTQCVNCSHFURGQECVKECRWKGLPREYVNGKRLDPCPHE
CPONSTETGSEADQCTACPHYKDSFPCVAPCPGKPDLSYMPWKYPDEEGMOQ
PCPINTHSCVDLDERGCPAEORASPATSIATVVGILLFLVGVVGVGILIKRRQKI
RKYTMLRLQLEVEPTSGAMPNOAOMRLKETELRKVKVYLGSGAFGTVYKGIWI
PDGENVKIPVAIKVLENTSPKANKELIDEAVMAGLSGSPVSVLLCICLSTVQLVT
QMPYGCGLDVRHGRGLSGDQLLNMCVQIAKMSYLEDVRLVHOLAARNVLKSP
NHVKTIDFGARLLDDETEYHADGGKVPKIAIALESILRRRTHQSDVMSYGVYWE
LMTFGAKYDIPAREIPDLLEGERLPQPICTIDYIMVMKCMIDSECRPRFREL
VSFSRMARDPQFVYQIONEDLGPSSPLDSTFYRSLLEDMDGDLVDAEVLVPOQGF
FFPDAPAGAGTAHRRHSSSTRSGGELTLGMEPSGEEPRSPRLAPSEGAGSDVEEG
ELMGATKQPSIDRDLPLQRYSEDTPLPTEITDGYVAPLACSPQPEYVNGEYR
PQPLPEGLPFPVRPAGATLERKLTSPGKNGVYKDVFTFGGAVENTPEYLVPRGSSA
SQHPFALCPAFNLXYDDQDPSESGSPNTFTGTPAENPEYLGLDVVPY"
87..3794
/mat_peptide
/genes="neu"
/product="p-185"
516..1988
/misc_feature
/genes="neu"
/notes="extracellular domain"
1989..2057
/misc_feature
/genes="neu"
/notes="putative; transmembrane domain"
2058..3794
/misc_feature
/genes="neu"
/notes="cytoplasmic domain"
845 a 1209 c 1182 g 826 t
BASE COUNT 845 a 1209 c 1182 g 826 t
ORIGIN
Query Match 1.4%; Score 22; DB 32; Length 4062;
Best Local Similarity 100.0%; Pred. No. 3.40e-02;
Matches 22; Conservative 0; Mismatches --0; Indels --0; Gaps 0;
Db 2524 AGGCATGAGCTACCTGGAGGA 2545
|||||
QY 1055 AGGCATGAGCTACCTGGAGGA 1076

RESULT 4
LOCUS DME132670 5423 bp DNA INV 09-FEB-1999
DEFINITION Drosophila melanogaster buttonless gene, exon 1.
ACCESSION A0132670
NID 94544137
VERSION A0132670.1 GI:4454137
KEYWORDS buttonless gene; homeodomain; 1(3)L4910; P-element.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
1 (bases 1 to 5423)
Zhang, Y.Q.
Direct Submission
Submitted (06-FEB-1999) Zhang Y.Q., Biology, University of Utah,
257 South 1400 East, Salt Lake City, Utah 84112-0840, USA
```

```

REFERENCE 2 (bases 1 to 5423)
AUTHORS Zhang, Y.Q. and Broadie, K.S.
TITLE Cloning, mapping and tissue-specific expression of Drosophila
clathrin-associated protein AP50 gene
JOURNAL Unpublished
FEATURES
Source Location/Qualifiers
1..5423
/organism="Drosophila melanogaster"
/notes="1(3)L4910"
/db_xref="taxon:7227"
/chromosome="3R"
/map="94B"
1..2346
/genes="buttonless"
1..2346
/genes="buttonless"
/number=1
1..2346
/genes="buttonless"
1786..22262
/genes="buttonless"
/notes="plamid rescue"
/codon_start=1
/product="homeodomain protein"
/protein_id="CAA10727.1"
/db_xref="PID:el386970"
/db_xref="PID:g4454138"
/db_xref="GI:4454138"
/translation="MKNSHANYQESCYEDITKSWLNQTEGYTDCSVYDHSASSY
ADYKLTNCTEANDQWQNEAPAGQHPQSRPKRAISSNRKRRTAFSKTOLKOLE
AEFCYSNLYLRRIEYTAVALTELRQYKVFQNRMRCKRIKLEEQGSSAKTDF"
5304..5311
/misc_feature
/notes="target recognition sequence"
BASE COUNT 1378 a 1267 c 1207 g 1571 t
ORIGIN
Query Match 1.4%; Score 21; DB 21; Length 5423;
Best Local Similarity 100.0%; Pred. No. 2.11e-01;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 121 GGTCTCTCGGCGTCTCTGCTGC 141
|||||
QY 1277 GGTCTCTCGGCGTCTCTGCTGC 1297

RESULT 5
LOCUS AC006494 162815 bp DNA HTG 30-MAR-1999
DEFINITION Drosophila melanogaster; Chromosome 3R; Region 94A12-B5; BAC clone
BACR48C17, WORKING DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION AC006494
NID 94544352
VERSION AC006494.3 GI:4544352
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
1 (bases 1 to 162815)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Khm, B., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclebb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 162815)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chavez, C., Chew, M., Ciesiolka, L.,
```

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomoton, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (03-FEB-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

**COMMENT**  
On Mar 30, 1999 this sequence version replaced gi:4454430. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 400 bases, phrap computed error rate <= 1/10.

\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1187: contig of 1187 bp in length  
1188 1267: gap of unknown length  
1268 3797: contig of 2530 bp in length  
3798 3877: gap of unknown length  
3878 7590: contig of 3713 bp in length  
7591 7670: gap of unknown length  
7671 14497: contig of 6827 bp in length  
14498 14577: gap of unknown length  
14578 22541: contig of 8064 bp in length  
22542 27221: gap of unknown length  
27222 37766: contig of 15045 bp in length  
37767 37846: gap of unknown length  
37847 54413: contig of 16567 bp in length  
54414 54493: gap of unknown length  
54494 77564: contig of 23071 bp in length  
77565 77644: gap of unknown length  
77645 162815: contig of 85171 bp in length.

# FEATURES

source

1. 162815  
/organism="Drosophila melanogaster"  
/strain="y2-cn-bw-sp"  
/db\_xref="taxon:7227"  
/clone="BACR48C17 (D504) RPCI-98 48.C.17"  
/chromosome="3"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"  
/map="94A12-B5"

**BASE COUNT** 45843 a 35560 c 35024 g 45746 t 642 others

## ORIGIN

Query Match 1.4%; Score 21; DB 20; Length 162815;  
Best Local Similarity 100.0%; Pred. No. 2,11e-01;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 22842 GGTCTTCGGCTCGCTGC 22862

QY 1277 GGTCTTCGGCTCGCTGC 1297

## RESULT

6 AC004929 183180 bp DNA HTG 12-JUN-1998  
LOCUS Homo sapiens clone D0917G04, WORKING DRAFT SEQUENCE, 35 unordered  
DEFINITION pieces.  
ACCESSION AC004929  
NID G3213068  
VERSION AC004929.1 GI:3213068  
KEYWORDS HTG; HTGS\_PHASE1.

## SOURCE

### ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominiidae; Homo.

### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

1 (bases 1 to 183180)  
Waterston, R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 183180)

Waterston, R.H.

Direct Submission

Submitted (12-JUN-1998) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1551: contig of 1551 bp in length  
1552 1570: gap of unknown length  
1571 3054: contig of 1484 bp in length  
3055 3073: gap of unknown length  
3074 4598: contig of 1525 bp in length  
4599 4617: gap of unknown length  
4618 6473: contig of 1835 bp in length  
6473 6491: gap of unknown length  
6492 8059: contig of 1567 bp in length  
8059 8077: gap of unknown length  
8078 9742: contig of 1665 bp in length  
9743 9761: gap of unknown length  
9762 12411: contig of 2630 bp in length  
12412 12430: gap of unknown length  
12431 14171: contig of 1741 bp in length  
14172 14190: gap of unknown length  
14191 15846: contig of 1656 bp in length  
15847 15865: gap of unknown length  
15866 17520: contig of 1655 bp in length  
17521 17539: gap of unknown length  
17540 19116: contig of 1577 bp in length  
19117 19135: gap of unknown length  
19136 20814: contig of 1679 bp in length  
20815 20833: gap of unknown length  
20834 22417: contig of 1584 bp in length  
22418 22436: gap of unknown length  
22437 24022: contig of 1586 bp in length  
24023 24041: gap of unknown length  
24042 25501: contig of 1460 bp in length  
25502 25520: gap of unknown length  
25521 27156: contig of 1636 bp in length  
27157 27175: gap of unknown length  
27176 28714: contig of 1539 bp in length  
28715 28733: gap of unknown length  
28734 30235: contig of 1502 bp in length  
30236 30254: gap of unknown length  
30255 32004: contig of 1750 bp in length  
32005 32023: gap of unknown length  
32024 34896: contig of 2873 bp in length  
34897 34915: gap of unknown length  
34916 38958: contig of 4043 bp in length  
38959 38977: gap of unknown length  
38978 44238: contig of 5261 bp in length  
44239 44257: gap of unknown length  
44258 55029: contig of 10772 bp in length  
55030 55048: gap of unknown length  
55049 66616: contig of 11568 bp in length  
66617 66635: gap of unknown length  
66636 86781: contig of 20146 bp in length  
86782 86800: gap of unknown length  
86801 107071: contig of 20271 bp in length

\* 107072 107090: gap of unknown length  
 \* 140184: contig of 33094 bp in length  
 \* 140185 140203: gap of unknown length  
 \* 140204 171124: contig of 30921 bp in length  
 \* 171125 171142: gap of unknown length  
 \* 171143 172670: contig of 1528 bp in length  
 \* 172671 172688: gap of unknown length  
 \* 172689 174207: contig of 1519 bp in length  
 \* 174208 174225: gap of unknown length  
 \* 174226 175954: contig of 1729 bp in length  
 \* 175955 175972: gap of unknown length  
 \* 175973 178202: contig of 2230 bp in length  
 \* 178203 178220: gap of unknown length  
 \* 178221 180109: contig of 1889 bp in length  
 \* 180110 180127: gap of unknown length  
 \* 180128 181704: contig of 1577 bp in length  
 \* 181705 181722: gap of unknown length  
 \* 181723 183180: contig of 1458 bp in length.

FEATURES  
 source  
 1..183180 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DJ0917G04"  
 BASE COUNT 48785 a 41389 c 42477 g 49890 t 639 others  
 ORIGIN

Query Match 1.4%; Score 22; DB 19; Length 183180;  
 Best Local Similarity 100.0%; Pred. No. 3.40e-02;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 51540 CCTCACCCAGCTTCCTCCCAA 51561

Cp 764 CCTCACCCAGCTTCCTCCCAA 743

RESULT 7 MUSPKV 211 bp mRNA ROD 12-AUG-1994  
 LOCUS Mus musculus protein tyrosine-kinase mRNA, partial cds.  
 DEFINITION L25762  
 ACCESSION  
 NID 9413748  
 VERSION L25762.1 GI:413748  
 KEYWORDS protein-tyrosine kinase.  
 SOURCE Mus musculus cDNA to mRNA.  
 ORGANISM Mus musculus  
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;  
 Murinae; Mus.

REFERENCE 1 (bases 1 to 211)  
 AUTHORS Hebert, B.  
 TITLE Identification de quatorze transcrits codant pour des protéines tyrosine kinases chez les cellules E-5; caractérisation partielle de trois nouveaux transcrits  
 JOURNAL Thesis (1993)  
 REFERENCE 2 (bases 1 to 211)  
 AUTHORS Hebert, B., Bergeron, J., Tijssen, P. and Potworowski, E. F.  
 TITLE Protein tyrosine kinases transcribed in a murine thymic medullary epithelial cell line  
 JOURNAL Gene 143, 257-260 (1994)  
 MEDLINE 94266162

FEATURES  
 source  
 1..211 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_line="E-5"  
 /cell\_type="murine thymic medullary epithelial cell line"  
 1..211  
 /gene="ptk"  
 <1..>211  
 /gene="ptk"  
 /note="Motifs VI thru IX of the catalytic domain.; putative"  
 /citation=[1]  
 /citation=[2]

/codon\_start=1  
 /function="novel member of src family of PTK"  
 /product="protein-tyrosine kinase"  
 /protein\_id="AAA40020.1"  
 /db\_xref="PID:G554265"  
 /db\_xref="GI:554265"  
 /translation="IDHRLDRAANILVGENLICKIADFGLARLIEDNEYTARQGAKE  
 PIKWTAPAEALYGRFTIKSDVWSFGI"  
 misc\_difference 55..58

/gene="PTK"  
 /note="Aside from wobble changes, this catalytic domain sequence has one codon change when compared to c-yes; putative"  
 /citation=[1]  
 /citation=[2]

BASE COUNT 67 a 41 c 46 g 57 t

Query Match 1.3%; Score 20; DB 32; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 184 AAGTCAGACGCTGTGTCCTT 203

Qy 1264 AAGTCAGACGCTGTGTCCTT 1283

RESULT 8 RERSVRC 1653 bp RNA VRL 01-JUN-1989  
 LOCUS Rous sarcoma virus Pra src gene (pp60).  
 DEFINITION X14718  
 ACCESSION  
 NID 961714  
 VERSION X14718.1 GI:61714  
 KEYWORDS src gene.  
 SOURCE Rous sarcoma virus.

ORGANISM Rous sarcoma virus.  
 Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 REFERENCE 1 (bases 1 to 1653)  
 AUTHORS Hackett, P. B.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAR-1989) Hackett P. B., University of Minnesota, Department of Genetics and Cell Biology, 250 Biosciences Centre, St Paul MN 55108-1095, U S A

REFERENCE 2 (bases 1 to 1653)  
 AUTHORS Liu, Z. J. and Hackett, P. B.  
 TITLE Sequence variation of the Rous sarcoma virus Pra src gene  
 JOURNAL Nucleic Acids Res. 17 (10), 3986 (1989)  
 MEDLINE 89282411  
 COMMENT \*map: 7054-8706.  
 FEATURES Location/Qualifiers  
 source  
 1..1653  
 /organism="Rous sarcoma virus"  
 /db\_xref="taxon:11886"

BASE COUNT 354 a 516 c 500 g 283 t

Query Match 1.3%; Score 20; DB 37; Length 1653;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+00;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1273 TGCAAGGTGGCTGACTTCGG 1292

Qy 1138 TGCAAGGTGGCTGACTTCGG 1157

RESULT 9 HUMHCKE 1926 bp mRNA PRI 08-NOV-1994  
 LOCUS Human hemopoietic cell protein-tyrosine kinase (HCK) gene, complete cds, clone HK24.  
 DEFINITION M16592  
 ACCESSION  
 NID g183913  
 VERSION M16592.1 GI:183913  
 KEYWORDS kinase; protein kinase; protein-tyrosine kinase.

SOURCE	Human mitogen-stimulated leukocyte, cDNA to mRNA, clone HK24.	
ORGANISM	Homo sapiens	gene
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 1926)	CDS
AUTHORS	Ziegler, S.F., Marth, J.D., Lewis, D.B. and Perlmutter, R.M.	
TITLE	Novel protein-tyrosine kinase gene (hck) preferentially expressed in cells of hematopoietic origin	
JOURNAL	Mol. Cell. Biol. 7 (6), 2276-2285 (1987)	
MEDLINE	87257943	
FEATURES	Location/Qualifiers	
source	1..1926	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/map="20q11-q12"	
mRNA	<1..1926	
gene	/note="HCK mRNA"	
	76..1593	
CDS	/gene="HCK"	
	76..1593	
	/gene="HCK"	
	/codon_start=1	
	/db_xref="GDB:G00-119-303"	
	/protein_id="AA52644.1"	
	/db_xref="PID:G306833"	
	/db_xref="GI:306833"	
	/translation="MGCMKSKFLOVGGNTFTSTASPHCPVYVDPDTSTIKPGPNS HNSNTPGIREAGSEDLIVVLYDYEAIIHEDLSFQGDQMVVLEESGEWKKARSLATR KEGYIPSNYARVDSLETEWFFKGISRDAERQLLAGNMLGSGMIRDSSETTKGSYS LSVRDYDPQGGTVKHYKIRTLNDGGFYISPRSTFTLOELVDHYKKGNDGLCKLSV PCMSKPOKPEKDAWEIPRESLKEKLGAGQFGEVMMATYKHTKVAVKTKPGSM SVEAFALANVMKTLQHDKLVKLHVVTKPIYIITEFMAKGLDFLKSDGSKOPL PKLIDFSQIAEGMAFISQRYNIHRDLRAANILVSASLVCKIADFGARVIEDNEYTA REGAFPIKWTAPAEINFGSTIKSDVWSFGILLMEIVTYGRIPYPGMNPVIRALE RGYRNPRENCEPEELYNIMRCWKNRPERPTEFYIQSVLDDFYTATESQYQQQP"	
BASE COUNT	497 a 522 c 520 g 387 t	
ORIGIN	1 bp upstream of EcoRI site; chromosome 20q11-q12.	
	Query Match 1.3%; Score 20; DB 29; Length 1926;	
	Best Local Similarity 100.0%; Pred. No. 1.23e+00;	
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	1321 AAGTCAGACGCTCTGGTCCTT 1340	
QY	1264 AAGTCAGACGCTCTGGTCCTT 1283	
RESULT	10	
LOCUS	HUMHCKA 2015 bp mRNA PRI 08-NOV-1994	
DEFINITION	Human hemopoietic cell protein-tyrosine kinase (HCK) gene, complete cds, clone lambda-a2/1a.	
ACCESSION	M16591	
NID	g183911	
VERSION	M16591.1 GI:183911	
KEYWORDS	kinase; protein kinase; protein-tyrosine kinase.	
SOURCE	Human hemopoietic cell, cDNA to mRNA, clone lambda-a2/1a.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 2015)	
AUTHORS	Quintrell, N., Lebo, R., Varmus, H., Bishop, J.M., Pettenati, M.J., Le Beau, M.M., Diaz, M.O. and Rowley, J.D.	
TITLE	Identification of a human gene (HCK) that encodes a protein-tyrosine kinase and is expressed in hemopoietic cells	
JOURNAL	Mol. Cell. Biol. 7 (6), 2267-2275 (1987)	
MEDLINE	87257942	
FEATURES	Location/Qualifiers	
source	1..2015	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/map="20q11-q12"	
mRNA	<1..2015	
	/note="HCK mRNA"	
	169..1686	
	/gene="HCK"	
	169..1686	
	/gene="HCK"	
	/codon_start=1	
	/db_xref="GDB:G00-119-303"	
	/protein_id="AA52643.1"	
	/db_xref="PID:G306832"	
	/db_xref="GI:306832"	
	/translation="MGSMKSKFLOVGGNTFTSTASPHCPVYVDPDTSTIKPGPNS HNSNTPGIREAGSEDLIVVLYDYEAIIHEDLSFQGDQMVVLEESGEWKKARSLATR KEGYIPSNYARVDSLETEWFFKGISRDAERQLLAGNMLGSGMIRDSSETTKGSYS LSVRDYDPQGGTVKHYKIRTLNDGGFYISPRSTFTLOELVDHYKKGNDGLCKLSV PCMSKPOKPEKDAWEIPRESLKEKLGAGQFGEVMMATYKHTKVAVKTKPGSM SVEAFALANVMKTLQHDKLVKLHVVTKPIYIITEFMAKGLDFLKSDGSKOPL PKLIDFSQIAEGMAFISQRYNIHRDLRAANILVSASLVCKIADFGARVIEDNEYTA REGAFPIKWTAPAEINFGSTIKSDVWSFGILLMEIVTYGRIPYPGMNPVIRALE RGYRNPRENCEPEELYNIMRCWKNRPERPTEFYIQSVLDDFYTATESQYQQQP"	
BASE COUNT	512 a 540 c 580 g 383 t	
ORIGIN	130 bp upstream of BamHI site; chromosome 20q11-q12.	
	Query Match 1.3%; Score 20; DB 29; Length 2015;	
	Best Local Similarity 100.0%; Pred. No. 1.23e+00;	
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	1414 AAGTCAGACGCTCTGGTCCTT 1433	
QY	1264 AAGTCAGACGCTCTGGTCCTT 1283	
RESULT	11	
LOCUS	HSHCKE11 2167 bp DNA PRI 16-JUL-1992	
DEFINITION	H. sapiens HCK gene for tyrosine kinase (PTK), exons 10-11.	
ACCESSION	X58742 X59742	
NID	g32043	
VERSION	X58742.1 GI:32043	
KEYWORDS	proto-oncogene; src family; T-cell receptor alpha-chain; Tyrosine kinase; V-alpha gene segment; variable region.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 2167)	
AUTHORS	Hradetzky, D.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-JUN-1991) D. Hradetzky, Chemotherapeutisches Forschungsinstitut, Georg-Speyer-Haus, Paul Ehrlich Str 42-44, 6000 Frankfurt 70, Federal Republic of Germany	
REFERENCE	2 (bases 1 to 2167)	
AUTHORS	Hradetzky, D., Streibhardt, K. and Rubsamen-Waigmann, H.	
TITLE	The genomic locus of the human hemopoietic-specific cell protein tyrosine kinase (PTK)-encoding gene (HCK) confirms conservation of exon-intron structure among human PTKs of the src family	
JOURNAL	Gene 113 (2), 275-280 (1992)	
MEDLINE	92241680	
COMMENT	See also X58736-X58740, X58744-X58769	
	See also X58741 and X58743.	
FEATURES	Location/Qualifiers	
source	1..2167	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="20"	
	/dev_stage="adult"	
	/tissue_type="spleen"	
	/clone_lib="genomic; TS48"	
	/clone="D640 H"	
	/map="q11-12"	
	<1..15	
	/number=9	
	16..1065	
	/gene="HCK"	
	intron	
	gene	

```

exon      16..169
/genes="HCK"
/number=10
/usedin=X58741:HCK_cds
/usedin=X58741:HCK_mRNA
intron    170..933
/genes="HCK"
/number=10
repeat_region 203..430
/notes="ALU repeat v"
934..1065
/genes="HCK"
/number=11
/usedin=X58741:HCK_cds
/usedin=X58741:HCK_mRNA
1066..>2167
/number=11
repeat_region 1281..1592
/notes="ALU repeat VI"
repeat_region 1835..2167
/notes="ALU repeat VII"
BASE COUNT 477 a 626 c 485 g 579 t
ORIGIN

```

```

Query Match      1.38; Score 20; DB 30; Length 2167;
Best Local Similarity 100.0%; Pred. No. 1.23e+00;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 996 AAGTCAGACGTCGTGGCTT 1015
|||||
QY 1264 AAGTCAGACGTCGTGGCTT 1283

```

```

RESULT 12
LOCUS      XHCFYN      2526 bp      mRNA      VRT      22-MAY-1991
DEFINITION Xiphophorus c-fyn (Xfyn) mRNA for protein p59(Xfyn).
ACCESSION  X54971
NID        964481
VERSION    X54971.1 GI:64481
KEYWORDS   fyn oncogene; oncogene.
SOURCE     Xiphophorus helleri.
ORGANISM   Xiphophorus helleri
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
            Atherinomorpha; Cyprinodontiformes; Cyprinodontoidae; Poeciliidae;
            Xiphophorus.
REFERENCE  1 (bases 1 to 2526)
AUTHORS   Schartl,M.
TITLE     Direct Submission
JOURNAL   Submitted (22-OCT-1990) Schartl M., Max-Planck-Institut fuer
            Biochemie, Genzentrum, Am Klopferspitz 18a, D-8033 Martinsried, FRG
REFERENCE  2 (bases 1 to 2526)
AUTHORS   Hannig,G., Ottlie,S. and Schartl,M.
TITLE     Conservation of structure and expression of the c-fyn and fyn genes
            in lower vertebrates
JOURNAL   Oncogene 6 (3), 361-369 (1991)
MEDLINE   91187435
FEATURES   Location/Qualifiers
            source
            1..2526
            /organism="Xiphophorus helleri"
            /strain="Rio Lancetilla"
            /isolate="Db-"
            /db_xref="taxon:8084"
            /clone="lambda 3911"
            1..>2526
            /notes="p59(Xfyn)"
            222..1835
            /genes="Xfyn"
            222..1835
            /genes="Xfyn"
            /codon_start=1
            /product="p59(Xfyn)"
            /protein_id="CAA38715.1"

```

```

/db_xref="PID:g64482"
/db_xref="GI:64482"
/translat="MGCVQCKDEATKLTDRDASISQAGYRYGADPTPQHPSFGV
TAIPNNFHPVGGVTVFGVNTSHTGTLRGGTGVTLFVALYDYEARTEDDLS
FRKGERFOILNSTEGDWDARSLTGGSGYIPSNVAPVDSIOAEDWYFGLGRKDAE
ROLLSTGNPRGTYLIRESETKGAFLSIRDWDEKGDHVKYIKRLDSSGYITTR
AOFDTLOQLVQHSYDRAAGCCRLVWFCHGMPRLADLSYTKDWWEIPRSLQLIKR
LGNGQFGEVWMTWNGTKVAVKTLKPGTSPESFLEEAQIMKKLRHDKLVOLYAVVS
EEPIYIYETMSKSLDLFLKDGEGRAKLPNLVDMAAQVAAAGYIEMMYIHRLR
SANLVGDNLVCKIADFLARLIEDNEVTAQGAFFPKWTAPAAALYGRITIKSDVW
SFGILLTELVTGGRVPYPGMNNREVLEQVERGMRPCPDQCPASLHMLQCKWKDPE
ERTFEYLQAFLEDYFTATEPQYQGNL"
polyA_site 674 a 673 c 707 g 472 t
BASE COUNT 674 a 673 c 707 g 472 t
ORIGIN
Query Match      1.38; Score 20; DB 24; Length 2526;
Best Local Similarity 100.0%; Pred. No. 1.23e+00;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1443 GACTTCGGCCTGGCGCGCT 1462
|||||
QY 1150 GACTTCGGCCTGGCGCGCT 1169

```

RESULT 13  
 LOCUS RSVPSRC 2545 bp RNA VRL 10-SEP-1992  
 DEFINITION Duck adapted Rous sarcoma virus (Pr-RSV-C) src gene and 3'-LTR.  
 ACCESSION X51861  
 NID 961896  
 VERSION X51861.1 GI:61896  
 KEYWORDS src gene.  
 SOURCE Rous sarcoma virus.  
 ORGANISM Rous sarcoma virus.  
 Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 REFERENCE 1 (bases 1 to 2545)  
 AUTHORS Zubak,S.V.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAR-1990) Zubak S., Institute Mol. Biol. & Genet.,  
 Academy of Science Ukr SSR, 252627 Kiev, Str Zabolotnogo 150, U S R  
 REFERENCE 2 (bases 1 to 2545)  
 AUTHORS Kashuba,V.I., Serge,z.V., Rynditch,A.V., Kavsan,V.M. and  
 Hlozanek,I.  
 JOURNAL Unpublished  
 FEATURES Location/Qualifiers  
 source  
 1..2545  
 /organism="Rous sarcoma virus"  
 /strain="Pr-RSV-C"  
 /isolate="da Pr-RSV-C"  
 /db\_xref="taxon:11896"  
 /cell\_line="Khaki Campbell duck embryo fibroblasts"  
 50..142  
 /rpt\_type=DIRECT  
 282..1862  
 /genes="scr"  
 282..1862  
 /genes="scr"  
 /codon\_start=1  
 /protein\_id="CAA36154.1"  
 /db\_xref="PID:g61897"  
 /db\_xref="GI:61897"  
 /db\_xref="SPIRMBL:Q07461"  
 /translat="MGSSKSRDPSQRRHSLEPPDSTHHGGFFASQTPDPAAPDAH  
 RNPSSFGTVATEPGKLEWNTSDTVSPQRAAGVTVFVALYDYESWTEDLSF  
 KGERLOIVNTEGYWMLAHLATGQGYIPSNVAPSDSIOAEWYFGKITRRESR  
 LLNPNPRGTELVKSETAKGAVCLSDPDKNAKGNVHKYIKLYSGGYITLRT  
 QFGSLQOLVAHYSKHKADGLCHRLNVCPTSPKQTOGLAKDAWEIPRSLRLEAKICG  
 CFGEVWMTWNGTTRVAIKTLKPGTNSPEAFLOEAQMKRFRHKLVLQLYAVSSEPI  
 YIVIEYMSKSLDLFLKDGEGRAKLPNLVDMAAQVAAAGYIEMMYIHRLR  
 LVGENLVCKVADFLARLIEDNEVTAQGAFFPKWTAPAAALYGRITIKSDVSGFI  
 LLTELTTKGRVPYPGMNNREVLEQVERGMRPCPECPESLHDLQCKQCKRDPERT

```
repeat_region FKYLQAOQLLPACVLEVAE"
1944..2043
/rpt_type=DIRECT
LTR 2212..2545
BASE COUNT 583 a 693 c 752 g 517 t
ORIGIN

Query Match 1.3%; Score 20; DB 37; Length 2545;
Best Local Similarity 100.0%; Pred. No. 1.23e+00;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1479 TCGAAGGTGGCTGACTCGG 1498
|||||
Qy 1138 TCGAAGGTGGCTGACTCGG 1157

RESULT 14
LOCUS DRAJ5030 2640 bp mRNA VRT 27-MAY-1998
DEFINITION Danio rerio mRNA for Eph-like receptor tyrosine kinase rtk4,
partial.
ACCESSION AJ005030
NID g3005932
VERSION AJ005030.1 GI:3005932
KEYWORDS Eph-like receptor tyrosine kinase.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinoidae; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 2640)
AUTHORS Cooke,J.E.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1998) Cooke J.E., University College London,
Department of Anatomy and Developmental Biology, Gower Street,
London, WC1E 6BT, UK
REFERENCE 2 (bases 1 to 2640)
AUTHORS Cooke,J.E., Xu,Q., Wilson,S.W. and Holder,N.
TITLE Characterisation of five novel zebrafish Eph-related receptor
tyrosine kinases suggests roles in patterning the neural plate
JOURNAL Dev. Genes Evol. 208: 515-531 (1997)
FEATURES
source Location/Qualifiers
1..2640
/organism="Danio rerio"
/db_xref="taxon:7955"
<1..>2640
/codon_start=1
/product="Eph-like receptor tyrosine kinase rtk4"
/protein_id="CAA08303.1"
/db_xref="PID:e1285052"
/db_xref="PID:g3005933"
/db_xref="GI:3005933"
/db_xref="SPTREMBL:O73879"
/translation="EVSMDEKNPIRTYQVCNVNWPSONWLRTHWTQGAQRIYI
EIKLRLDCNSLPVGIGKTFNLYLESDDNERYAHESRFSKIDITVADESFTQV
DIGRIMKLTENVRGDGLVSRAGFYLAQDVAGIALVSFVFKKPLAVRNLAQFP
DTVAGDGSLSVEGVSCVDHSEOEVPKMGADGWLPIGNCLCPNPGVEERNEQC
QVCKGVYRSLSSDVGCPCPLHSYSVKEGATSCDCKGYVRSSTDPAAECTPPSA
PHNLISWNETHSVLLDWSFPLSSGGRLDITNVYCKQCVRTQRCSPGDDVRSQPQ
LSLRSTRVSVHQAHTNPTFIQWAVNGSVKHNFSPEQAVSVLTITNQAASFMSVNG
SKDITRHTLALFDOPEKPNPILLEYEVKYEKQNSRYRIVRTTSNADIKDITPL
TSYVHFVARTAGYAGDFSAPEFTSTNTVAAPVGVGMSSAVLLLLVAGCVVLLIL
IITITIKRRSKYKTKOGDVQVRIYVDPTEDPNOAIRFAKELNTSICKI
EKVIGYFPGVGCGRLLPKGRREICVAIKTLKAGFTQQRDLFSEASVIGQDPHN
LIHLEGVYTKCPWILTEYMEVNSLDMFLKNCRCFTVYIQLVGLIRIASGMKYLSD
MSYVHRDAARNLNVNLCVKVDFGMSRVLEDEPEGATPIRTGKIPIRWTAPEAIT
YRKFTASDVSYSYGVMEVMSYGERPYWDMNSQDVKAIEEGRVLPMPPECPLAHQ
LMLECMREADRPKFSQIVNMKLIKNPNTLKRATAGGASRSHPSSTINHPSECSAP
SLASVDDWLKLGLEQYENFNTA"
BASE COUNT 675 a 661 c 719 g 585 t
ORIGIN

repeat_region 1944..2043
/rpt_type=DIRECT
LTR 2212..2545
BASE COUNT 583 a 693 c 752 g 517 t
ORIGIN

Query Match 1.3%; Score 20; DB 24; Length 2640;
Best Local Similarity 100.0%; Pred. No. 1.23e+00;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2394 GTCATGCTGGAGTCTGGA 2413
|||||
Qy 1419 GTCATGCTGGAGTCTGGA 1438

RESULT 15
LOCUS RERSVH19 3123 bp RNA ROD 30-MAY-1996
DEFINITION Hamster H-19 proviral DNA (LTR- v-src -LTR).
ACCESSION X15345
NID 961706
VERSION X15345.1 GI:61706
KEYWORDS long terminal repeat; oncogene; provirus; src oncogene; src
oncogene viral.
SOURCE golden hamster.
ORGANISM Mesocricetus auratus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
REFERENCE 1 (bases 1 to 3123)
AUTHORS Bodor,J.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1989) Bodor J., Institute of Molecular Genetics
Czechoslovak Academy of Sciences, Flemingovo 2, 166 37 Prague,
Czechoslovakia
REFERENCE 2 (bases 1 to 3123)
AUTHORS Bodor,J., Poliak,E., Pichrtova,J., Geryk,J. and Svoboda,J.
TITLE Complete nucleotide sequence of LTR- v-src, LTR provirus H-19
JOURNAL Nucleic Acids Res. 17 (21), 8869 (1989)
MEDLINE 90067864
COMMENT The H-19 proviral seq was generated in the hamster genome by src
mRNA reverse transcription and integration.
FEATURES
source Location/Qualifiers
1..3123
/organism="Mesocricetus auratus"
/db_xref="taxon:10036"
1..3123
/organism="Rous sarcoma virus"
/db_xref="taxon:11886"
<1..>120
/note="hamster DNA"
/organism="Mesocricetus auratus"
/db_xref="taxon:10036"
/haplotype="21-22 hypodiploid"
/tissue_type="mesenchymal"
/cell_type="sarcoma cells induced by XC rescued PR-RSV-C"
/cell_line="H19"
/clone_lib="lambda NM1149"
/clone="H19clC2.1"
121..3097
/organism="Rous sarcoma virus"
/db_xref="taxon:11886"
121..126
/note="direct repeat 1"
127..459
/note="5' LTR"
329..335
/note="put. TATA-box"
352..357
/note="pot. polyA signal"
359..379
/note="direct repeat 2 (R)"
756..2992
/note="exon (v-src)"
831..2411
/note="pp60v-src (AA 1-526)"
/codon_start=1
/protein_id="CAA33404.1"
/db_xref="PID:g61707"
/db_xref="GI:61707"
/translation="MGSSKSKPKDPSQRRSLPPDTHHGGFPASQTPDETAAPDAH"
```



```
RNPSRSFGTVAPEPKLFMGFNISDITVTSQPRAGALAGGVTTFVALYDYESWTETDLSF
KGERLOIVNTEGDMWLAHSITGOTGYIPSNYVAPSDSIOAEEMVFGKITPRESER
LLLNENPRGTEFLVRKSETAKGAYCLSVSDEDNAGPNYKHXYKICKLYSGGFYITST
QFGSLOQLVAYISKHADGCHRLTNVCPSTKFTQGLAKDAWEIPRESLRLEAKLGG
CEGEVWMTNWTTRVAIKTLKPGTMSPEAFLOEAQVMKKLRHEKLDVLYAVVSEEP
YVIEYMSKSLDLEFKGEMGYLRLPOLVDMAAQIASGMAYVERNNYVHRDLRAANI
LVGENLVKRVADFGARLJEDNEYTAROGAKFPKWTAPEALYGRFTIKSDWVSFGI
LLETLTGKRVFPGMVNREVLDQVERAYRMFCPPCESLHDLMCQCQCRKDEPERT
FKYLOAQLLPACVLEVAE"
2760..3091
/note="3' LTR"
promoter
2964..2970
/note="put. TATA-box"
misc_feature
2986..2991
/note="pot. polyA signal"
repeat_region
2993..3013
/note="direct repeat 2 (R)"
repeat_region
3092..3097
/note="direct repeat 1"
3098..>3123
/note="hamster DNA"
/organism="Mesocricetus auratus"
/db_xref="taxon:10036"
/haplotype="21-22 hypodiploid"
/tissue_type="mesenchymal"
/cell_type="sarcoma cells induced by XC rescued PR-RSV-C"
/cell_line="H19"
/clone_lib="lambda NM1149"
/clone="H19clC2.1"
BASE COUNT 716 a 843 c 936 g 628 t
ORIGIN
```

```
Query Match 1.38; Score 20; DB 32; Length 3123;
Best Local Similarity 100.0%; Pred.No. 1.23e+00;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2028 TGCAGGTGGCTGACTCGG 2047
QY 1138 TGCAGGTGGCTGACTCGG 1157
|||||
```

Search completed: Thu May 20 18:37:08 1999  
Job time : 7753 secs.

\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu May 20 19:34:52 1999; MasPar time 411.24 Seconds  
806.566 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-099-053-1  
Description: (1-1548) from US09090533.seq  
Perfect Score: 1548  
N.A. Sequence: 1 GCTCGCGGCTCCATGCGC.....CCAACGGCTCTGGGCTCCAGC 1548  
Comp: CGACGCGCGGAGGTACCGG.....GGTTCCGAGACCGGAGGTGC

Scoring table: TABLE jmetric  
Gap 60

Mismatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-gene35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 7.892; Variance 3.155; scale 2.502

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1548	100.0	1548	59	Human SAD encoding cD	0.00e+00
2	30	1.9	30	59	Human SAD probe 5557.	4.57e-09
3	21	1.4	21	59	Human SAD PCR primer	3.96e-02
4	21	1.4	21	59	Human SAD PCR primer	3.96e-02
5	21	1.4	21	59	Human SAD PCR primer	3.96e-02
6	21	1.4	21	59	Human SAD PCR primer	3.96e-02
7	21	1.4	21	59	Human SAD PCR primer	3.96e-02
8	20	1.3	20	59	Human SAD PCR primer	1.99e-01
9	18	1.2	18	59	Human SAD PCR primer	4.37e+00

C	10	18	1.2	18	59	V81768	Human SAD PCR primer	4.37e+00
C	11	18	1.2	18	59	V81761	Human SAD RACE primer	4.37e+00
C	12	19	1.2	19	59	V81760	Human SAD RACE primer	9.54e-01
C	13	18	1.2	19	52	V69680	T84X gene specific ri	4.37e+00
C	14	18	1.2	255	40	V20456	Human c-src1l oncogen	4.37e+00
C	15	18	1.2	656	24	T13754	Partial sequence of e	4.37e+00
C	16	19	1.2	1177	38	T89400	Methods for diagnosis	9.54e-01
C	17	18	1.2	1602	7	Q46687	Chicken pp60 c-src ge	4.37e+00
C	18	18	1.2	1611	7	Q46688	Human pp60 c-src gene	4.37e+00
C	19	19	1.2	1628	53	V71969	Mouse ActRIIB4 recept	9.54e-01
C	20	19	1.2	1651	53	V71967	Mouse ActRIIB2 recept	9.54e-01
C	21	19	1.2	1699	53	V71968	Mouse ActRIIB3 recept	9.54e-01
C	22	19	1.2	1722	53	V71966	Mouse ActRIIB1 recept	9.54e-01
C	23	18	1.2	2507	13	Q81189	Breast tumour kinase,	4.37e+00
C	24	18	1.2	2774	30	T60434	Human intracellular t	4.37e+00
C	25	18	1.2	2789	30	T60433	Human intracellular t	4.37e+00
C	26	18	1.2	3836	13	Q73652	HSV-2 strain 25766 in	4.37e+00
C	27	18	1.2	3836	52	V64102	HSV-2 strain 25766 gH	4.37e+00
C	28	18	1.2	4031	33	T84485	Mouse alpha-1 collage	4.37e+00
C	29	18	1.2	4049	15	Q90660	Eph-related PTK Cex5.	4.37e+00
C	30	18	1.2	4097	15	Q90657	Eph-related PTK Cex5+	4.37e+00
C	31	18	1.2	4868	42	V26068	Human canalicular mul	4.37e+00
C	32	18	1.2	5586	37	T94023	Human canalicular mul	4.37e+00
C	33	18	1.2	9439	52	V69284	Human variably charge	4.37e+00
C	34	18	1.2	20387	50	V62159	HSV-2 strain SB5 Cont	4.37e+00
C	35	18	1.2	26338	49	V62134	HSV-2 strain SB5 Cont	4.37e+00
C	36	18	1.2	117213	51	V62176	HSV-2 strain SB5 Cont	4.37e+00
C	37	17	1.1	17	59	V81764	Human SAD PCR primer	1.90e+01
C	38	17	1.1	17	59	V81767	Human SAD PCR primer	1.90e+01
C	39	17	1.1	1458	55	V08951	Human ATG-1639 protei	1.90e+01
C	40	17	1.1	1640	27	T45868	Human bone morphogeni	1.90e+01
C	41	17	1.1	1767	59	V80657	Human FRAZ2LED protei	1.90e+01
C	42	17	1.1	1969	59	V80658	Partial human FRAZ2LE	1.90e+01
C	43	17	1.1	3268	8	Q49164	113 kD ISGF-3alpha ge	1.90e+01
C	44	17	1.1	3370	13	Q79534	Bovine tracheal antim	1.90e+01
C	45	17	1.1	4989	24	Q99245	Insulin like growth f	1.90e+01

ALIGNMENTS

RESULT	1
ID	V81743 standard; cDNA; 1548 BP.
AC	V81743;
DT	10-MAR-1999 (first entry)
DE	Human SAD encoding cDNA.
KW	PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW	type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW	neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW	Parkinson's disease; Huntington's disease; ss.
OS	Homo sapiens.
FH	Key
FT	CDS
FT	49..1515
FT	Location/Qualifiers
FT	/*tag- a
PN	WO9849317-A2.
PD	05-NOV-1998.
PF	27-APR-1998; U08439.
PR	23-OCT-1997; US-063595.
PR	28-APR-1997; US-044428.
PR	20-MAY-1997; US-047222.
PR	11-JUN-1997; US-049477.
PR	11-JUN-1997; US-049756.
PR	18-JUN-1997; US-049914.
PA	(SUGC-) SUGEN INC.
PI	App H, Clary D, Courtneidge SA, Hui TH, Jallal B,
PI	Markby D, Onrust S, Peles E, Plozman GD;
DR	WPI; 99-009434/01.
DR	P-PSDB; W89248.
PT	New nucleic acid encoding specific protein tyrosine phosphatases -
PT	useful for identifying specific modulators for treatment and
PT	prevention of cancer and neurodegenerative disease
PS	Claim 2; Page 146; 193pp; English.
CC	The present invention describes isolated, enriched or purified nucleic
CC	acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The



PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
 PI Markby D, Onrust S, Peles E, Plowman GD;  
 DR WPI: 99-009434/01.  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Example 6; Page 87; 193pp; English.  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
 CC (PTPs) and are used to identify substances that modulate their activity  
 CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
 CC These substances are used to treat or prevent diseases associated with  
 CC abnormal signal transduction pathways that involve the proteins,  
 CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
 CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
 CC promote neuronal survival, particularly for treating Alzheimer's,  
 CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
 CC polynucleotides encoding the proteins can be used as probes to identify  
 CC and clone related sequences; to detect protein-encoded RNA; to generate  
 CC transgenic animals and in gene therapy (optionally after mutation). Ab  
 CC are used to determine the proteins. The present sequence represents a  
 CC probe for human SAD.  
 SQ Sequence 30 BP; 5 A; 11 C; 8 G; 6 T;  
 Query Match 1.9%; Score 30; DB 59; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 4.57e-09;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 tggagcggccacactcgaatcgccctt 30  
 QY 715 TGGAGCGGCCACACTCCGAATCGCCCTT 744  
 RESULT 3  
 ID V81769 standard; DNA; 21 BP.  
 AC V81769;  
 DT 10-MAR-1999 (first entry)  
 DE Human SAD PCR primer 6642.  
 KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9849317-A2.  
 PD 05-NOV-1998.  
 PF 27-APR-1998; U08439.  
 PR 23-OCT-1997; US-063595.  
 PR 28-APR-1997; US-044428.  
 PR 20-MAY-1997; US-047222.  
 PR 11-JUN-1997; US-049477.  
 PR 11-JUN-1997; US-049756.  
 PR 18-JUN-1997; US-049914.  
 PA (SUGEN-) SUGEN INC.  
 PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
 PI Markby D, Onrust S, Peles E, Plowman GD;  
 DR WPI: 99-009434/01.  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Example 6; Page 88; 193pp; English.  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
 CC (PTPs) and are used to identify substances that modulate their activity  
 CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
 CC These substances are used to treat or prevent diseases associated with  
 CC abnormal signal transduction pathways that involve the proteins,  
 CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
 CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
 CC promote neuronal survival, particularly for treating Alzheimer's,  
 CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the

CC polynucleotides encoding the proteins can be used as probes to identify  
 CC and clone related sequences; to detect protein-encoded RNA; to generate  
 CC transgenic animals and in gene therapy (optionally after mutation). Ab  
 CC are used to determine the proteins. The present sequence represents a  
 CC PCR primer for human SAD.  
 SQ Sequence 21 BP; 4 A; 5 C; 8 G; 4 T;  
 Query Match 1.4%; Score 21; DB 59; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.96e-02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 atgagccgcttcctcagagg 21  
 QY 49 ATGGAGCGCTTCCTCAGAGG 69  
 RESULT 4  
 ID V81771 standard; DNA; 21 BP.  
 AC V81771;  
 DT 10-MAR-1999 (first entry)  
 DE Human SAD PCR primer 6643.  
 KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9849317-A2.  
 PD 05-NOV-1998.  
 PF 27-APR-1998; U08439.  
 PR 23-OCT-1997; US-063595.  
 PR 28-APR-1997; US-044428.  
 PR 20-MAY-1997; US-047222.  
 PR 11-JUN-1997; US-049477.  
 PR 11-JUN-1997; US-049756.  
 PR 18-JUN-1997; US-049914.  
 PA (SUGEN-) SUGEN INC.  
 PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
 PI Markby D, Onrust S, Peles E, Plowman GD;  
 DR WPI: 99-009434/01.  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Example 6; Page 88; 193pp; English.  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
 CC (PTPs) and are used to identify substances that modulate their activity  
 CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
 CC These substances are used to treat or prevent diseases associated with  
 CC abnormal signal transduction pathways that involve the proteins,  
 CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
 CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
 CC promote neuronal survival, particularly for treating Alzheimer's,  
 CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
 CC polynucleotides encoding the proteins can be used as probes to identify  
 CC and clone related sequences; to detect protein-encoded RNA; to generate  
 CC transgenic animals and in gene therapy (optionally after mutation). Ab  
 CC are used to determine the proteins. The present sequence represents a  
 CC PCR primer for human SAD.  
 SQ Sequence 21 BP; 6 A; 6 C; 6 G; 3 T;  
 Query Match 1.4%; Score 21; DB 59; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.96e-02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 aggcacactggaagctgattcc 21  
 QY 641 AGGCCAACTGGAGCTGATCC 661  
 RESULT 5  
 ID V81773 standard; DNA; 21 BP.

AC V81773;  
 DT 10-MAR-1999 (first entry)  
 DE Human SAD PCR primer 5284.  
 KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9849317-A2.  
 PD 05-NOV-1998.  
 PF 27-APR-1998; U08439.  
 PR 23-OCT-1997; US-063595.  
 PR 28-APR-1997; US-044428.  
 PR 20-MAY-1997; US-047222.  
 PR 11-JUN-1997; US-049477.  
 PR 11-JUN-1997; US-049756.  
 PR 18-JUN-1997; US-049914.  
 PA (SUGEN) SUGEN INC.  
 PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
 PI Markby D, Onrust S, Peles E, Plowman GD;  
 DR WPI; 99-009434/01.  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Example 7; Page 90; 193pp; English.  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
 CC (PTPs) and are used to identify substances that modulate their activity  
 CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
 CC These substances are used to treat or prevent diseases associated with  
 CC abnormal signal transduction pathways that involve the proteins,  
 CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
 CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
 CC promote neuronal survival, particularly for treating Alzheimer's,  
 CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
 CC polynucleotides encoding the proteins can be used as probes to identify  
 CC and clone related sequences; to detect protein-encoded RNA; to generate  
 CC transgenic animals and in gene therapy (optionally after mutation). Ab  
 CC are used to determine the proteins. The present sequence represents a  
 CC PCR primer for human SAD.  
 SQ Sequence 21 BP; 7 A; 7 C; 5 G; 2 T;

Query Match 1.4%; Score 21; DB 59; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.96e-02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tcgccaaggagatccagacac 21  
 |||||||||||||||||||||  
 Qy 857 TCGCCAGGAGATCCAGACAC 877

RESULT 6  
 ID V81774 standard; DNA; 21 BP.  
 AC V81774;  
 DT 10-MAR-1999 (first entry)  
 DE Human SAD PCR primer 5285.  
 KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9849317-A2.  
 PD 05-NOV-1998.  
 PF 27-APR-1998; U08439.  
 PR 23-OCT-1997; US-063595.  
 PR 28-APR-1997; US-044428.  
 PR 20-MAY-1997; US-047222.  
 PR 11-JUN-1997; US-049477.  
 PR 11-JUN-1997; US-049756.  
 PR 18-JUN-1997; US-049914.  
 PA (SUGEN) SUGEN INC.  
 PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
 PI Markby D, Onrust S, Peles E, Plowman GD;  
 DR WPI; 99-009434/01.  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Example 7; Page 90; 193pp; English.  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
 CC (PTPs) and are used to identify substances that modulate their activity  
 CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
 CC These substances are used to treat or prevent diseases associated with  
 CC abnormal signal transduction pathways that involve the proteins,  
 CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
 CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
 CC promote neuronal survival, particularly for treating Alzheimer's,  
 CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
 CC polynucleotides encoding the proteins can be used as probes to identify  
 CC and clone related sequences; to detect protein-encoded RNA; to generate  
 CC transgenic animals and in gene therapy (optionally after mutation). Ab  
 CC are used to determine the proteins. The present sequence represents a  
 CC PCR primer for human SAD.  
 SQ Sequence 21 BP; 7 A; 7 C; 5 G; 2 T;

Query Match 1.4%; Score 21; DB 59; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.96e-02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tcgccaaggagatccagacac 21  
 |||||||||||||||||||||  
 Qy 857 TCGCCAGGAGATCCAGACAC 877

RESULT 7  
 ID V81770 standard; DNA; 21 BP.  
 AC V81770;  
 DT 10-MAR-1999 (first entry)  
 DE Human SAD PCR primer 6644.  
 KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9849317-A2.  
 PD 05-NOV-1998.  
 PF 27-APR-1998; U08439.  
 PR 23-OCT-1997; US-063595.  
 PR 28-APR-1997; US-044428.  
 PR 20-MAY-1997; US-047222.  
 PR 11-JUN-1997; US-049477.  
 PR 11-JUN-1997; US-049756.  
 PR 18-JUN-1997; US-049914.  
 PA (SUGEN) SUGEN INC.  
 PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
 PI Markby D, Onrust S, Peles E, Plowman GD;  
 DR WPI; 99-009434/01.  
 PT New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease  
 PS Example 6; Page 88; 193pp; English.  
 CC The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
 CC (PTPs) and are used to identify substances that modulate their activity  
 CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
 CC These substances are used to treat or prevent diseases associated with  
 CC abnormal signal transduction pathways that involve the proteins,  
 CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
 CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
 CC promote neuronal survival, particularly for treating Alzheimer's,  
 CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
 CC polynucleotides encoding the proteins can be used as probes to identify  
 CC and clone related sequences; to detect protein-encoded RNA; to generate  
 CC transgenic animals and in gene therapy (optionally after mutation). Ab  
 CC are used to determine the proteins. The present sequence represents a  
 CC PCR primer for human SAD.  
 SQ Sequence 21 BP; 5 A; 7 C; 6 G; 3 T;

Query Match 1.4%; Score 21; DB 59; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.96e-02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gaagtcagccacctgcaggc 21  
 |||||||||||||||||||||  
 Cp 1155 GAAGTCAGCCACCTGCAGGC 1135

CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
CC polynucleotides encoding the proteins can be used as probes to identify  
CC and clone related sequences; to detect protein-encoded RNA; to generate  
CC transgenic animals and in gene therapy (optionally after mutation). Ab  
CC are used to determine the proteins. The present sequence represents a  
CC PCR primer for human SAD.  
SQ Sequence 21 BP; 4 A; 10 C; 3 G; 4 T;

Query Match 1.4%; Score 21; DB 59; Length 21;  
Best Local Similarity 100.0%; Pred. No. 3.96e-02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tcaccagcttctcccaag 21

Cp 761 TCACCAGCTTCTCCCAAG 741

## RESULT 8

ID V81772 standard; DNA; 20 BP.  
AC V81772;  
DT 10-MAR-1999 (first entry)  
DE Human SAD PCR primer 6645.  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGEN) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plowman GD;  
DR WPI: 99-009434/01.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Example 6; Page 88; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
CC (PTPs) and are used to identify substances that modulate their activity  
CC (i.e. agonists and antagonists) including NBP in vivo or in vitro.  
CC These substances are used to treat or prevent diseases associated with  
CC abnormal signal transduction pathways that involve the proteins,  
CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
CC promote neuronal survival, particularly for treating Alzheimer's,  
CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
CC polynucleotides encoding the proteins can be used as probes to identify  
CC and clone related sequences; to detect protein-encoded RNA; to generate  
CC transgenic animals and in gene therapy (optionally after mutation). Ab  
CC are used to determine the proteins. The present sequence represents a  
CC PCR primer for human SAD.  
SQ Sequence 20 BP; 3 A; 5 C; 9 G; 3 T;

Query Match 1.3%; Score 20; DB 59; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.99e-01;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gctgagccagcagcgttgg 20

Cp 1548 GCTGAGCCAGCAGCCTGG 1529

## RESULT 9

ID V81765 standard; DNA; 18 BP.  
AC V81765;  
DT 10-MAR-1999 (first entry)  
DE Human SAD PCR primer 5554.  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGEN) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plowman GD;  
DR WPI: 99-009434/01.

PT New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Example 6; Page 86; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
CC (PTPs) and are used to identify substances that modulate their activity  
CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
CC These substances are used to treat or prevent diseases associated with  
CC abnormal signal transduction pathways that involve the proteins,  
CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
CC promote neuronal survival, particularly for treating Alzheimer's,  
CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
CC polynucleotides encoding the proteins can be used as probes to identify  
CC and clone related sequences; to detect protein-encoded RNA; to generate  
CC transgenic animals and in gene therapy (optionally after mutation). Ab  
CC are used to determine the proteins. The present sequence represents a  
CC PCR primer for human SAD.  
SQ Sequence 18 BP; 6 A; 9 C; 1 G; 2 T;

Query Match 1.2%; Score 18; DB 59; Length 18;  
Best Local Similarity 100.0%; Pred. No. 4.37e+00;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 ccacacctccccaaagta 18

Cp 783 CCACACCTCCCAAGTA 766

## RESULT 10

ID V81768 standard; DNA; 18 BP.  
AC V81768;  
DT 10-MAR-1999 (first entry)  
DE Human SAD PCR primer 6119.  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.

PR 18-JUN-1997; US-049914.  
PA (SUGEN-) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plowman GD;  
PI WPI; 99-009434/01.  
DR New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Example 6; Page 87; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
CC (PTPs) and are used to identify substances that modulate their activity  
CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
CC These substances are used to treat or prevent diseases associated with  
CC abnormal signal transduction pathways that involve the proteins,  
CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
CC promote neuronal survival, particularly for treating Alzheimer's,  
CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
CC polynucleotides encoding the proteins can be used as probes to identify  
CC and clone related sequences; to detect protein-encoded RNA; to generate  
CC transgenic animals and in gene therapy (optionally after mutation). Ab  
CC are used to determine the proteins. The present sequence represents a  
CC PCR primer for human SAD.  
SQ Sequence 18 BP; 4 A; 6 C; 6 G; 2 T;  
  
Query Match 1.2%; Score 18; DB 59; Length 18;  
Best Local Similarity 100.0%; Pred. No. 4.37e+00;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 1 cttcgagggcacagcgc 18  
|||||  
Cp 295 CTTGAGGGGCACAGGCC 278  
  
RESULT 11  
ID V81761 standard; DNA; 18 BP.  
AC V81761;  
DE 10-MAR-1999 (first entry)  
DE Human SAD RACE primer 5848  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN W09849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGEN-) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plowman GD;  
PI WPI; 99-009434/01.  
DR New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Example 6; Page 86; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
CC (PTPs) and are used to identify substances that modulate their activity  
CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
CC These substances are used to treat or prevent diseases associated with  
CC abnormal signal transduction pathways that involve the proteins,  
CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to

CC promote neuronal survival, particularly for treating Alzheimer's,  
CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
CC polynucleotides encoding the proteins can be used as probes to identify  
CC and clone related sequences; to detect protein-encoded RNA; to generate  
CC transgenic animals and in gene therapy (optionally after mutation). Ab  
CC are used to determine the proteins. The present sequence represents a  
CC RACE primer for human SAD.  
SQ Sequence 18 BP; 4 A; 4 C; 7 G; 3 T;  
  
Query Match 1.2%; Score 18; DB 59; Length 18;  
Best Local Similarity 100.0%; Pred. No. 4.37e+00;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 1 ggtagaggctgcacagc 18  
|||||  
Cp 589 GGTAGAGGCTGCCATCAG 572  
  
RESULT 12  
ID V81760 standard; DNA; 19 BP.  
AC V81760;  
DE 10-MAR-1999 (first entry)  
DE Human SAD RACE primer 5556.  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN W09849317-A2.  
PD 05-NOV-1998.  
PF 27-APR-1998; U08439.  
PR 23-OCT-1997; US-063595.  
PR 28-APR-1997; US-044428.  
PR 20-MAY-1997; US-047222.  
PR 11-JUN-1997; US-049477.  
PR 11-JUN-1997; US-049756.  
PR 18-JUN-1997; US-049914.  
PA (SUGEN-) SUGEN INC.  
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B,  
PI Markby D, Onrust S, Peles E, Plowman GD;  
PI WPI; 99-009434/01.  
DR New nucleic acid encoding specific protein tyrosine phosphatases -  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS Example 6; Page 86; 193pp; English.  
CC The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC above proteins, other than ALK-7, are protein tyrosine phosphatases  
CC (PTPs) and are used to identify substances that modulate their activity  
CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.  
CC These substances are used to treat or prevent diseases associated with  
CC abnormal signal transduction pathways that involve the proteins,  
CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of  
CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to  
CC promote neuronal survival, particularly for treating Alzheimer's,  
CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the  
CC polynucleotides encoding the proteins can be used as probes to identify  
CC and clone related sequences; to detect protein-encoded RNA; to generate  
CC transgenic animals and in gene therapy (optionally after mutation). Ab  
CC are used to determine the proteins. The present sequence represents a  
CC RACE primer for human SAD.  
SQ Sequence 19 BP; 3 A; 3 C; 6 G; 7 T;  
  
Query Match 1.2%; Score 19; DB 59; Length 19;  
Best Local Similarity 100.0%; Pred. No. 9.54e-01;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 1 agtgaagcttcattgtggct 19  
|||||  
Cp 852 ACTGAGCTTCATGTTGGCT 834

```
RESULT 13
ID V69680 standard; DNA; 19 BP.
AC V69680, 1999 (first entry)
DE TB4X gene specific right primer.
KW Non-recombining region; human; Y chromosome; X homologue; testis; sperm;
KW infertility; gene alteration; inhibitor; TB4X; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9846747-A2.
PD 22-OCT-1998.
PR 10-APR-1998; U07115.
PR 11-APR-1997; US-041877.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Lann BR, Page DC.
DR WPI; 98-568729/48.
PT Novel genes in the non-combining region of Y chromosome - useful to
PT diagnose if male infertility or reduced sperm count has a genetic
PT basis.
PS Example 2; Page 30; 54pp; English.
PS Sequences V69651 to V69684 represent PCR primers used for localisation
CC of the genes of the invention which occur on the non-recombining region
CC of the human Y chromosome. The gene sequences fall into two classes: (1)
CC X-homologous DNA which are expressed in many organs, having functional
CC X homologues and (2) testis-specific DNA sequences. Y chromosomal DNA
CC from males with known conditions such as infertility and reduced sperm
CC count can be assessed using the invention to determine whether the
CC condition is associated with or caused by the occurrence of the gene or
CC gene alteration. Candidate inhibitors in vitro assays. 4 T;
CC genes can be assessed using in vitro assays. 4 T;
SQ Sequence 19 BP; 4 A; 8 C; 3 G; 4 T;

Query Match 1.2%; Score 18; DB 52; Length 19;
Best Local Similarity 100.0%; Pred. No. 4.37e+00;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 ctccccaagtagccttc 19
Cp 777 CTCCCCAAGTAGCCTTC 760

RESULT 14
ID V20456 standard; DNA; 255 BP.
AC V20456;
DE 17-JUN-1998 (first entry)
KW Human c-src11 oncogene.
KW Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
KW cancer; antisense oligonucleotide; c-src11; ds.
OS Homo sapiens.
PN US5734039-A.
PD 31-MAR-1998.
PR 15-SEP-1994; 306691.
PR 15-SEP-1994; US-306691.
PA (UJJE-) UNIV JEFFERSON THOMAS.
PI Calabretta B, Skorski T;
DR WPI; 98-229882/20.
PT Anticancer composition comprising two anti-sense oligo:nucleotide(s)
PT - targeting cytoplasmic and nuclear oncogene(s)
PS Claim 1; Column 105-106; 9pp; English.
CC The present sequence represents an oncogene from the present invention.
CC The present invention describes a composition which comprises two
CC antisense oligonucleotides. The first oligonucleotide is specific for a
CC cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
CC c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC c-rel, c-vav, c-ski, c-spi, cyclin D1, PML/RAR alpha, AML1/MTG8,
CC E2A/p1 and ALL-1/AF-4. The composition is used for treating cancer.
CC The combination of antisense oligonucleotides has synergistically
CC enhanced ability to inhibit growth of cancer cells.
SQ Sequence 255 BP; 44 A; 86 C; 86 G; 39 T;

Query Match 1.2%; Score 18; DB 40; Length 255;
```

```
Best Local Similarity 100.0%; Pred. No. 4.37e+00;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 148 acctgcagcctctctgg 165
QY 974 ACCTGCAGGCCTTCCTGG 991

RESULT 15
ID T13754 standard; DNA; 656 BP.
AC T13754;
DE 17-DEC-1996 (first entry)
DE Partial sequence of endoglin.
KW Endoglin; TGF-beta; beta-glycan; receptor; diagnosis; detection;
KW gene therapy; haemorrhagic telangiectasia; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT intron 1..296
FT intron /*tag= a
FT /*label= Intron 10.
FT exon 297..554
FT /*tag= b
FT /*label= Exon 11.
FT intron 555..656
FT /*tag= c
FT /*label= Intron 11.
PN WO9616975-A1.
PD 06-JUN-1996.
PR 29-NOV-1995; U15428.
PR 29-NOV-1994; US-346129.
PA (HSCR-) HSC RES & DEV LP.
PA (UYDU-) UNIV DUKE.
PI Letarte M, Marchuk DA, McAllister K;
DR WPI; 96-286827/29.
PT Human gene for endoglin (transforming growth factor beta binding
PT protein) - useful in diagnosis and gene therapy of hereditary
PT haemorrhagic telangiectasia
PS Claim 6; Page 41; 71pp; English.
CC Oligonucleotides derived from introns of the endoglin gene can be
CC used as primers for amplifying a single exon of the endoglin gene
CC for its use in diagnosis of haemorrhagic telangiectasia (HHT). DNA
CC encoding endoglin can be used for gene therapy of HHT which is
CC caused by inheritance of a defective gene, e.g. endoglin, beta-
CC glycan, TGF-beta type I or II receptor or TGF-beta/activin type I
CC receptor.
SQ Sequence 656 BP; 132 A; 196 C; 202 G; 126 T;

Query Match 1.2%; Score 18; DB 24; Length 656;
Best Local Similarity 100.0%; Pred. No. 4.37e+00;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 188 gtgaggactcagggtgg 205
Cp 1523 GTGAGGACTCAGGGGTGG 1506

Search completed: Thu May 20 19:49:17 1999
Job time : 865 secs.
```



\*\*\*\*\*

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu May 20 18:37:31 1999; MasPar time 2885.34 Seconds  
Tabular output not generated. 1257:120 Million 'cell' updates/sec

Title: >US-099-053-1  
Description: (1-1548) from US0909053.seq  
Perfect Score: 1548  
N.A. Sequence: 1 GCTCGCGGCTCCCATGCC.....CCAAAGCTCTGGGTCCAGC 1548  
Comp: CGAGCGCGGAGGTACCG.....GGTTCGAGACCGGAGGTCCG

Scoring table: TABLE jmetric  
Gap 60

Mismatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
6:em\_est9 7:em\_gss1  
Database: genbank-est11  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_est38 38:gb\_gss2  
39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6  
Statistics: Mean 9.358; Variance 1.389; scale 6.737

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	72	4.7	332 26	A1373274 qz48f07.x1 NCI_CGAP_Ki	1.67e-123
2	22	1.4	249 33	W17139 zb12c04.r1 Soares_brea	5.49e-09
3	20	1.3	196 35	AA035407 zk26h10.r1 Soares_feta	1.13e-05
c 4	20	1.3	417 20	AA090391 ak13g03.s1 Soares_preg	1.13e-05
5	20	1.3	467 9	AA149096 z141g09.r1 Soares_preg	1.13e-05
c 6	20	1.3	493 41	AQ374666 RPC111-158L24.TJ RPCI	1.13e-05
c 7	20	1.3	602 41	AQ347709 RPC111-137G11.TV RPCI	1.13e-05
8	19	1.2	165 37	B03480 CSRL-17ell-u csRL flow	4.14e-04
c 9	18	1.2	167 22	AI059300 UI-R-Cl-1f-a-11-0-UI-s	1.28e-02
10	18	1.2	341 31	D60912 HUM137H10B Clontech hu	1.28e-02

11	19	1.2	364 25	A1282729 q156d10.x1 NCI_CGAP_Es	4.14e-04
12	18	1.2	370 30	R73590 Y197B06.r1 Soares_brea	1.28e-02
c 13	18	1.2	387 18	AA814869 C006c08.s1 NCI_CGAP_GC	1.28e-02
c 14	18	1.2	387 38	AA054911 CIT-HSP-234N3.TR CIT-	1.28e-02
c 15	19	1.2	400 22	A1058611 UI-R-Cl-Ky-b-06-0-UI-s	4.14e-04
c 16	18	1.2	409 32	H79782 yu10a02.s1 Soares_feta	1.28e-02
17	18	1.2	411 30	H14982 ym19g11.s1 Soares_infa	1.28e-02
18	18	1.2	419 16	AA677204 z160f03.s1 Soares_feta	1.28e-02
19	18	1.2	430 40	AQ185427 HS_2247.B1.G04_MF CIT	1.28e-02
20	19	1.2	433 16	AA616543 v010f03.r1 Barstead mo	4.14e-04
c 21	19	1.2	435 26	AI402163 GH09173.3prime GH Dros	4.14e-04
c 22	19	1.2	439 11	AA345776 EST51859 Gall bladder	4.14e-04
c 23	19	1.2	448 8	T54433 y006e04.r2 Stratagene	4.14e-04
c 24	18	1.2	449 23	A1159821 qc78g10.x1 Soares_plac	1.28e-02
c 25	18	1.2	450 11	AA309011 EST19795 Colon carcin	1.28e-02
c 26	18	1.2	455 32	H86410 Y993d07.r1 Soares_reti	1.28e-02
c 27	19	1.2	460 19	R12164 Y48e06.r1 Soares_infa	4.14e-04
c 28	18	1.2	469 30	H09004 Y196g01.r1 Soares_infa	1.28e-02
c 29	19	1.2	474 22	AI071532 UI-R-Cl-Ky-b-06-0-UI-s	4.14e-04
30	18	1.2	482 13	AA449875 zx37c05.r1 Soares_mous	4.14e-04
31	18	1.2	485 31	H40255 YP59c12.r1 Soares_feta	1.28e-02
c 32	18	1.2	521 42	AQ421412 RPCI-11-168F19.TJ RPCI	1.28e-02
c 33	19	1.2	546 36	AA080464 mm88g08.r1 Stratagene	4.14e-04
34	18	1.2	546 10	AA224533 z115g12.r1 Stratagene	1.28e-02
35	19	1.2	564 23	A1154821 ud80a02.r1 Soares_mous	4.14e-04
36	18	1.2	567 42	AQ427364 CITBI-E1-2563M14.TF CI	1.28e-02
37	19	1.2	582 18	AA822071 vp23h07.r1 Stratagene	4.14e-04
c 38	19	1.2	594 36	AA073075 mm79a03.r1 Stratagene	4.14e-04
c 39	18	1.2	601 37	FR0017364 F_rubripes CSS sequenc	1.28e-02
c 40	19	1.2	606 9	AA150506 z108g08.s1 Soares_preg	4.14e-04
c 41	19	1.2	637 9	AA203536 zx56h10.r1 Soares_feta	4.14e-04
c 42	19	1.2	645 24	A1193645 qv71f11.x1 Soares_feta	4.14e-04
c 43	19	1.2	676 16	AA645871 vs36b06.r1 Stratagene	4.14e-04
c 44	18	1.2	716 24	AI221122 qg91a12.x1 Soares_NFL	1.28e-02
c 45	19	1.2	821 28	AU050975 AU050975 Sugano mouse	4.14e-04

ALIGNMENTS

1 A1373274 332 bp mRNA EST 16-FEB-1999  
LOCUS qz48f07.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2030149 3'  
DEFINITION similar to SW:SRM\_MOUSE Q62270 TYROSINE-PROTEIN KINASE SRM ; mRNA  
sequence.  
ACCESSION A1373274  
NID 94153140  
VERSION A1373274.1 GI:4153140  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 332)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jun 15, 1998 this sequence version replaced gi:3223249.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 434 Std Error: 0.00  
Seq primer: -40UP from Gibco.

```

FEATURES
Source
1. .332 Location/Qualifiers
/organism="Homo sapiens"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="IMAGE:2030149"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
BASE COUNT 59 a 115 c 86 g 72 t
ORIGIN
Query Match 4.7%; Score 72; DB 26; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.67e-123;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 261 AGTCGGGGCCAGCCAAAGTCTCCACTACCGGGTCTCATGGCAGCTGATGCAGCCT 320
|||||
QY 525 AGTCGGGGCCAGCCAAAGTCTCCACTACCGGGTCTCATGGCAGCTGATGCAGCCT 584
|||||
Db 321 CTACCTGCAGAA 332
QY 585 CTACCTGCAGAA 596
|||||

RESULT 2
LOCUS W17139 249 bp mRNA EST 29-APR-1996
DEFINITION zb12c04.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:301830 5' similar to gb:M64347 FIBROBLAST GROWTH FACTOR
RECEPTOR 3 PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION W17139
NID g1291518
VERSION W17139.1 GI:1291518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 249)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On Nov 29, 1993 this sequence version replaced gi:634607.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: mob.REGA+ET
High quality sequence stop: 1.
FEATURES
Source
1. .249 Location/Qualifiers
/organism="Homo sapiens"
/notes="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5,
AACTGGAAGAATTTCGGCGCGCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/db_xref="GDB:3757717"
[5'-TCTTACCAATCTCAAGTGGGAGCGCGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."
/db_xref="GDB:1246761"
/db_xref="taxon:9606"
/clone="IMAGE:301830"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 46 a 48 c 65 g 44 t
ORIGIN
Query Match 1.4%; Score 22; DB 33; Length 249;
Best Local Similarity 88.0%; Pred. No. 5.49e-09;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db 81 ACGTCTGGTCTCTTNGNGTCTCTGCT 105
|||||
QY 1271 ACGTCTGGTCTCTTNGNGTCTCTGCT 1295
|||||

RESULT 3
LOCUS AA035407 196 bp mRNA EST 10-MAY-1997
DEFINITION zK26h10.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:471715 5' similar to gb:M16591 TYROSINE-PROTEIN KINASE HCK
(HUMAN);, mRNA sequence.
ACCESSION AA035407
NID g1507064
VERSION AA035407.1 GI:1507064
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 196)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)
On Apr 5, 1995 this sequence version replaced gi:760783.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 774 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 183.
FEATURES
Source
1. .196 Location/Qualifiers
/organism="Homo sapiens"
/notes="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5,
AACTGGAAGAATTTCGGCGCGCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/db_xref="GDB:3757717"

```

```
/db_xref="taxon:9606"
/map="703C03:1"
/clone="IMAGE:471715"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/lab_host="DH10B"

BASE COUNT      42 a  57 c  53 g  44 t
ORIGIN

Query Match      1.3%; Score 20; DB 35; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.13e-05;
Matches          20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 117 AAGTCAGAGCTGTGTCCTT 136
      |||||||
Qy 1264 AAGTCAGAGCTGTGTCCTT 1283

RESULT 4
LOCUS      AA890391      417 bp      mRNA      EST      04-JAN-1999
DEFINITION aki303.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            IMAGE:1405876 3', mRNA sequence.
ACCESSION  AA890391
NID         93017270
VERSION     93017270
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 417)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
JOURNAL     Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:140564.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html

Insert Length: 1009 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 401.
Location/Qualifiers
1. .417
/organism="Homo sapiens"
/notes="Organ: parathyroid gland; Vector: pT7m3D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5'-
TGTACCAATCTGAAGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTTTT
T-3']", double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
/db_xref="taxon:9606"
/map="13p"
/clone="IMAGE:1405876"
```

```
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/lab_host="DH10B"

BASE COUNT      122 a  96 c  106 g  93 t
ORIGIN

Query Match      1.3%; Score 20; DB 20; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.13e-05;
Matches          20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 349 AGCCACGCCACAGCCTTCC 368
      |||||||
Cp 802 AGCCACGCCACAGCCTTCC 783

RESULT 5
LOCUS      RA149096      467 bp      mRNA      EST      19-MAY-1997
DEFINITION z141909.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:504544 5' similar to gb:M16591 TYROSINE-PROTEIN KINASE HCK
            (HUMAN);, mRNA sequence.
ACCESSION  RA149096
NID         g1719549
VERSION     RA149096.1 GI:1719549
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 467)
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E.,
            Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
            WashU-Merck EST Project
            Unpublished (1995)
JOURNAL     Unpublished (1995)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407090.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 748 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 381.
Location/Qualifiers
1. .467
/organism="Homo sapiens"
/notes="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGTGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/db_xref="GDB:3809532"
/db_xref="taxon:9606"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/clone="IMAGE:504544"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/lab_host="DH10B"

BASE COUNT      110 a  140 c  133 g  83 t  1 others
ORIGIN

Query Match      1.3%; Score 20; DB 9; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.13e-05;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 91 AAGTCAGAGCTGTGTCCTT 110  
 |||||  
 Qy 1264 AAGTCAGAGCTGTGTCCTT 1283

RESULT 6  
 LOCUS AQ374666 493 bp DNA GSS 05-MAR-1999  
 DEFINITION RPC111-158L24.TJ RPC111 Homo sapiens genomic clone R-158L24,  
 genomic survey sequence.

ACCESSION AQ374666  
 NID 94345889  
 VERSION AQ374666.1 GI:4345689  
 KEYWORDS GSS.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 493)  
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and  
 Venter,J.C.

## REFERENCE

1 (bases 1 to 493)  
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and  
 Venter,J.C.  
 Use of BAC End Sequences from Library RPC111 for Sequence-Ready Map

## TITLE

## JOURNAL

## COMMENT

Unpublished (1997)  
 Other\_GSSs: RPC111-158L24.TV

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

For clone availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

Research Genetics (info@resgen.com). BAC end search page:

[http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html).

Seq primer: SP6

Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..493

/organism="Homo sapiens"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

/db\_xref="GDB:7560575"

/db\_xref="taxon:9606"

/clone="R-158L24"

/clone\_lib="RPC111"

/sex="Male"

/cell\_type="Lymphocytes"

BASE COUNT 155 a 125 c 111 g 101 t 1 others

## ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 1.13e-05;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 AGCCGAGCCAGGCGCTTCC 144

|||||

Cp 802 AGCCGAGCCAGGCGCTTCC 783

## RESULT

## LOCUS

AQ347709 602 bp DNA GSS 21-JAN-1999

DEFINITION RPC111-137G11.TV RPC111 Homo sapiens genomic clone R-137G11,

genomic survey sequence.

ACCESSION AQ347709

NID 94172605

VERSION AQ347709.1 GI:4172605

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

For clone availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

Research Genetics (info@resgen.com). BAC end search page:

[http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html).

Seq primer: T7

Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..602

/organism="Homo sapiens"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

/db\_xref="GDB:7552378"

/db\_xref="taxon:9606"

/clone="R-137G11"

/clone\_lib="RPC111"

/sex="Male"

/cell\_type="Lymphocytes"

BASE COUNT 196 a 144 c 137 g 124 t 1 others

## ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 1.13e-05;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 137 AGCCGAGCCAGGCGCTTCC 156

|||||

Cp 802 AGCCGAGCCAGGCGCTTCC 783

## RESULT

## LOCUS

B03480 165 bp DNA GSS 26-JUN-1996

DEFINITION CSRL-17ell-u CSRL flow sorted Chromosome 11 specific cosmid Homo

sapiens genomic clone CSRL-17ell, genomic survey sequence.

ACCESSION B03480

NID 91412758

VERSION B03480.1 GI:1412758

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 165)

Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,

Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S.,

Harris,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M.,

Kupfer,K. and Garner,H.R.

Genomic Sequence Sampled Map of Chromosome 11

Unpublished (1996)

Contact: Evans GA, Shane Probst

Mcdermott Center for Human Growth and Development

University of Texas Southwestern Medical Center At Dallas

5323 Harry Hines Blvd, Dallas TX 75235-8591

Tel: 214-648-1600

Fax: 214-648-1666

Email: gevas@utsw.swmed.edu, shane@mcdermott.swmed.edu

Seq primer: T7

Class: cosmid ends

High quality sequence stop: 165...

# FEATURES

Location/Qualifiers  
1..165  
/organism="Homo sapiens"  
/note="Vector: sCos-1; Human Chromosome 11 specific cosmid library prepared from flow sorted human Chromosome 11 derived from Chinese Hamster Ovary (CHO) monochromosomal somatic cell hybrid, J1"  
/db\_xref="taxon:9606"  
/clone="CSR1-17ell"  
/clone\_lib="CSR1 flow sorted Chromosome.11 specific cosmid"  
/sex="female"  
/cell\_type="chimeric hamster somatic cell hybrid"  
50 a 37 c 41 g 28 t 9 others

# BASE COUNT

## ORIGIN

Query Match 1.2%; Score 19; DB 37; Length 165;  
Best Local Similarity 100.0%; Pred.No. 4.14e-04;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 134 CAAGGCCAACTGGAGCTG 152  
QY 639 CAAGGCCAACTGGAGCTG 657  
|||||

# RESULT

LOCUS AI059300 167 bp mRNA EST 11-FEB-1999  
DEFINITION UI-R-C1-1f-a-11-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone  
ACCESSION AI059300  
NID 93333077  
VERSION AI059300.1 GI:3333077  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 167)  
REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.  
AUTHORS Normalization and subtraction: two approaches to facilitate gene discovery  
TITLE  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797766.

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to identify it as a clone from the normalized adult 12-pay-embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics The following repetitive elements were found in this cDNA sequence: 1-61, >POLY\_A\$imple\_repeat  
Seq primer: M13 Forward.

# FEATURES

## source

Location/Qualifiers  
1..167  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-C1 library is a subtracted library derived from the UI-R-CO library, which is a subtracted library derived from the UI-R-Al and UI-R-El libraries. The UI-R-Al library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-El library consisted of a mixture of

individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3'5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-CO clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-CO library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

/db\_xref="taxon:10116"  
/clone="UI-R-C1-1f-a-11-0-UI"  
/clone\_lib="UI-R-C1"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
BASE COUNT 24 a 31 c 34 g 78 t  
ORIGIN

Query Match 1.2%; Score 18; DB 22; Length 167;  
Best Local Similarity 100.0%; Pred.No. 1.28e-02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 76 ACTGGGTCAGTGTGGG 93

Cp 170 ACTGGGTCAGTGTGGG 153

# RESULT

LOCUS D60912 341 bp mRNA EST 28-AUG-1995  
DEFINITION HUM137H10B Clontech human fetal brain polyA+ mRNA (#5535) Homo sapiens cDNA clone GEN-137H10 5', mRNA sequence.  
ACCESSION D60912  
NID 9962551  
VERSION D60912.1 GI:962551  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 341)  
REFERENCE Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.  
Fujiwara et al. (1995)  
TITLE Unpublished (1995)  
JOURNAL On May 18, 1995 this sequence version replaced gi:811044.  
COMMENT

Contact: Tsutomu Fujiwara  
Otsuka GEN Research Institute  
Otsuka Pharmaceutical Co.,Ltd  
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035  
Insert Length: 342 Std Error: 0.00  
High quality sequence stop: 276.

# FEATURES

## source

Location/Qualifiers  
1..341  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GEN-137H10"  
/clone\_lib="Clontech human fetal brain polyA+ mRNA (#5535)"

# BASE COUNT

## ORIGIN

48 a 110 c 125 g 57 t 1 others

```

Query Match      1.2%; Score 18; DB 31; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.28e-02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 130 CCGGCGCTGGAGGAGCTG 147
Qy 610 CCGGCGCTGGAGGAGCTG 627

RESULT 11
LOCUS      AI282729      364 bp      mRNA      EST      21-DEC-1998
DEFINITION qt56d10.x1 NCI_CGAP_Eso2 Homo sapiens cDNA clone IMAGE:1960147 3'
similar to SW:MS2_HUMAN P78325 CELL SURFACE ANTIGEN MS2 PRECURSOR
; contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION  AI282729
NID         G3920962
VERSION     AI282729.1 GI:3920962
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 364)
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     On May 10, 1995 this sequence version replaced gi:806203.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 613
High quality sequence stops: 329
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 329.
Location/Qualifiers
1..370
/organism="Homo sapiens"
/note="Organ: breast; Vector: pT7T3b (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pT7T3 vector (Pharmacia).
Library went through one round of normalization to a Cot =
230. Library constructed by Bento Soares and M. Fatima
Bonaldo."
/db_xref="GDB:570370"
/db_xref="taxon:9606"
/clone="IMAGE:156659"
/clone_lib="Soares breast 2NbHBst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      89 a 107 c 96 g
ORIGIN
Query Match      1.2%; Score 18; DB 30; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.28e-02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 324 TCTTCTGGGACAGATCT 341
Qy 89 TCTTCTGGGACAGATCT 106

RESULT 13
LOCUS      AA814869      387 bp      mRNA      EST      13-FEB-1998
DEFINITION oc06c08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1340078 3',
mrna sequence.
ACCESSION  AA814869
NID         G2884465
VERSION     AA814869.1 GI:2884465
KEYWORDS    EST.
SOURCE      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 387)
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     On May 10, 1995 this sequence version replaced gi:806203.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 613
High quality sequence stops: 329
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 329.
Location/Qualifiers
1..370
/organism="Homo sapiens"
/note="Organ: breast; Vector: pT7T3b (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pT7T3 vector (Pharmacia).
Library went through one round of normalization to a Cot =
230. Library constructed by Bento Soares and M. Fatima
Bonaldo."
/db_xref="GDB:570370"
/db_xref="taxon:9606"
/clone="IMAGE:156659"
/clone_lib="Soares breast 2NbHBst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      89 a 107 c 96 g
ORIGIN
Query Match      1.2%; Score 19; DB 25; Length 364;
Best Local Similarity 100.0%; Pred. No. 4.14e-04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 90 CCGGCGGCTGCTGAGG 108
Qy 1040 CCGGCGGCTGCTGAGG 1058

RESULT 12
LOCUS      R73690      370 bp      mRNA      EST      05-JUN-1995
DEFINITION Y797b06.r1 Soares breast 2NbHBst Homo sapiens cDNA clone
IMAGE:156659 5', mRNA sequence.
ACCESSION  R73690
NID         G2884465
VERSION     R73690.1 GI:2884465
KEYWORDS    EST.
SOURCE      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 370)
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     On May 10, 1995 this sequence version replaced gi:806203.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 613
High quality sequence stops: 329
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 329.
Location/Qualifiers
1..370
/organism="Homo sapiens"
/note="Organ: breast; Vector: pT7T3b (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pT7T3 vector (Pharmacia).
Library went through one round of normalization to a Cot =
230. Library constructed by Bento Soares and M. Fatima
Bonaldo."
/db_xref="GDB:570370"
/db_xref="taxon:9606"
/clone="IMAGE:156659"
/clone_lib="Soares breast 2NbHBst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      89 a 107 c 96 g
ORIGIN
Query Match      1.2%; Score 19; DB 25; Length 364;
Best Local Similarity 100.0%; Pred. No. 4.14e-04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 90 CCGGCGGCTGCTGAGG 108
Qy 1040 CCGGCGGCTGCTGAGG 1058

RESULT 12
LOCUS      R73690      370 bp      mRNA      EST      05-JUN-1995
DEFINITION Y797b06.r1 Soares breast 2NbHBst Homo sapiens cDNA clone
IMAGE:156659 5', mRNA sequence.
ACCESSION  R73690
NID         G2884465
VERSION     R73690.1 GI:2884465
KEYWORDS    EST.
SOURCE      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 370)
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     On May 10, 1995 this sequence version replaced gi:806203.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 613
High quality sequence stops: 329
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 329.
Location/Qualifiers
1..370
/organism="Homo sapiens"
/note="Organ: breast; Vector: pT7T3b (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pT7T3 vector (Pharmacia).
Library went through one round of normalization to a Cot =
230. Library constructed by Bento Soares and M. Fatima
Bonaldo."
/db_xref="GDB:570370"
/db_xref="taxon:9606"
/clone="IMAGE:156659"
/clone_lib="Soares breast 2NbHBst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      89 a 107 c 96 g
ORIGIN
Query Match      1.2%; Score 19; DB 25; Length 364;
Best Local Similarity 100.0%; Pred. No. 4.14e-04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 90 CCGGCGGCTGCTGAGG 108
Qy 1040 CCGGCGGCTGCTGAGG 1058

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797230.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldino, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 327.
Location/Qualifiers
1. .387
/organism="Homo sapiens"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CEER). cDNA synthesis was
primed with a Not I- oligo(dT) primer
[5'-TGTACCAATCGAAGTGGAGCGCCGCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldino."
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 92 a 112 c 93 g 90 t
ORIGIN
Query Match 1.2%; Score 18; DB 18; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.28e-02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 303 CTCGCCAAGTAGCCTTC 320
|||||
Cp 777 CTCGCCAAGTAGCCTTC 760

RESULT 14
LOCUS AQ054911 387 bp DNA GSS 29-JUL-1998
DEFINITION CIT-HSP-2343N3.TR CIT-HSP Homo sapiens genomic clone 2343N3,
genomic survey sequence.
ACCESSION AQ054911
NID 93351517
VERSION AQ054911.1 GI:3351517
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS Adams,M.D., Rounsley S.D., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,

Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)
Other_GSSs: CIT-HSP-2343N3.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamads@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1. .387
/organism="Homo sapiens"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
/db_xref="taxon:9606"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
BASE COUNT 80 a 118 c 121 g 68 t
ORIGIN
Query Match 1.2%; Score 18; DB 38; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.28e-02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 219 CATGCCCTCAGCCACCTG 236
|||||
Cp 1062 CATGCCCTCAGCCACCTG 1045

RESULT 15
LOCUS AI058611 400 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C1-ky-b-06-0-UI.s1 UI-R-C1 Rattus norvegicus CDNA clone
UI-R-C1-ky-b-06-0-UI 3', mRNA sequence.
ACCESSION AI058611
NID 93323388
VERSION AI058611.1 GI:33323388
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 14, 1998 this sequence version replaced gi:2043323.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult Ovary library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics
Seq primer: M13 Forward.
Location/Qualifiers

```

source

```

1. 400
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C1) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
/db_xref="taxon:10116"
/map="933H10: 3; 3q12.1-3q13.32; 17q21"
/clone="UI-R-C1-ky-b-06-0-UI"
/clone_lib="UI-R-C1"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"

BASE COUNT      105 a      89 c      94 g      112 t
ORIGIN

```

```

Query Match      1.2%; Score 19; DB 22; Length 400;
Best Local Similarity 100.0%; Pred.No. 4.14e-04;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 186 TGACTTGATGACCTTGATC 204
    |||||
Cp 834 TGACTTGATGACCTTGATC 816

```

Search completed: Thu May 20 19:34:29 1999  
Job time : 3418 secs.